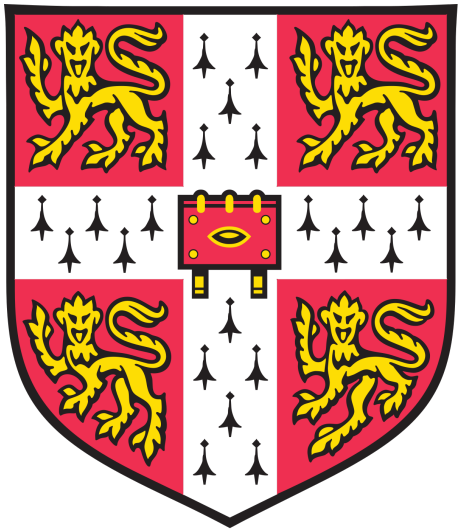
Integrative Analysis of the Human Gut Phageome Using a Metagenomics Approach



Luis Fernando Camarillo Guerrero

Gonville & Caius College
University of Cambridge

This thesis is submitted for the degree of
Doctor of Philosophy

August 2020

Dedicated to my parents (Rosa María and Leopoldo) and sister (Marisol) for loving me and always supporting me in an unconditional manner from Day 1 of my life

Dedicada a mis padres (Leopoldo y Rosa María) y hermana (Marisol) por quererme y apoyarme de una manera incondicional desde mi primer día de vida.

Declaration

This thesis is the result of my own work and includes nothing which is the outcome of work done in collaboration except as declared in the Preface and specified in the text. It is not substantially the same as any that I have submitted, or, is being concurrently submitted for a degree or diploma or other qualification at the University of Cambridge or any other University or similar institution except as declared in the Preface and specified in the text. I further state that no substantial part of my thesis has already been submitted, or, is being concurrently submitted for any such degree, diploma or other qualification at the University of Cambridge or any other University or similar institution except as declared in the Preface and specified in the text. It does not exceed the prescribed word limit for the relevant Degree Committee

Luis Fernando Camarillo Guerrero
August 2020

Summary

Integrative Analysis of the Human Gut Phageome Using a Metagenomics Approach

Luis Fernando Camarillo Guerrero

Bacteriophages (or phages; viruses that infect bacteria and archaea) profoundly influence microbial communities. Given the impact of the gut microbiome composition and function on human health, there is a growing focus on phages that inhabit the gut ecosystem. However, the extent of viral diversity, biology, and worldwide epidemiology of gut phages remain largely unknown. In this thesis, I carry out a comprehensive genomic analysis of gut phages by harnessing the biggest collection of phage genomes, gut bacteria isolates, and human gut metagenomes.

I begin by introducing the Gut Phage Database (GPD) which is the largest genomic resource to date of human gut phage genomes and product of mining 28,060 faecal metagenomes and 2898 gut bacteria isolate genomes. I use machine learning to improve the quality of the predictions and investigate ways to organise the viral diversity in order to improve the characterisation of gut phages in downstream analyses.

Afterwards, I describe common functions and auxiliary metabolic genes encoded by human gut phages. I also highlight instances of hypervariable domains which may indicate the presence of phage receptor binding proteins. I then shift the focus to the analysis of two clades of gut phages, namely the Gubaphage and the *Picovirinae* subfamily. The Gubaphage is a novel phage clade uncovered in this work which is highly prevalent across the world. The *Picovirinae* clade was the most common predicted phage taxonomy in GPD. Host assignment allows me to study patterns of phage diversity across bacterial clades of the human gut and investigate their host range.

Finally, I analyse global patterns of the human gut phageome and its association with lifestyle and bacterial composition. I assess the idea of a core virome as well as in what degree my data agrees with this concept.

Acknowledgements

I would like to thank my supervisor Trevor Lawley for trusting me and allowing me to embark on a PhD project of my own. I'm aware of how fortunate I am to have had this freedom. Trevor, I am very grateful for your continuous guidance, and support throughout my PhD journey.

I would like to thank my colleagues from my lab team162 who I have been fortunate to learn a lot from, in particular Yan Shao and Hilary Browne. Yan not only became a great friend, he challenged me throughout my PhD with stimulating discussions about my project. Thanks Yan, I appreciate all your support and you intellectually pushed me to reach new heights in my knowledge about phages. Hilary, I appreciate your accessibility and the time you invested proofreading presentations/reports/thesis/manuscript, you give amazing advice.

I also would like to thank my manuscript collaborator Alexandre Almeida. Alex, your positive outlook was always refreshing. I learned a lot working along with someone of your scientific calibre. Thanks for all your advice and friendship.

I extend my gratitude to the Wellcome Sanger Institute for funding my PhD, Carl Anderson, Christina Hedberg-Delouka, Annabel Smith, and the Pathogen Informatics team.

I would like to acknowledge Darwin College which was the foundation of my social life in Cambridge. To all the people I met there, I appreciate your friendship and have fond memories of you. To the Darwinian Michael Schneider, who actually I met at Imperial College London, thanks for your friendship and for all the nights out at the Cambridge Bops. Honourable mention to the Darwin Salsa society, every week I was looking forward to go dancing on Monday at 8:00 pm. I would also like to thank the Mexicans living in Cambridge, I had so much fun hanging out with you all and really made me feel at home.

I thank Jose Manuel Aguilar Yañez for playing an integral role in my formation as a Scientist and always believing in me.

Last but not least, I thank my family to whom I dedicate this thesis. Your unconditional love and support were vital for me throughout this endeavour – *¡Lo logramos!*

Publications

Camarillo-Guerrero L.F., Almeida A., Rangel-Pineros G., Finn R.D., Lawley T. (2020). Massive expansion of human gut bacteriophage diversity. (In press). *Cell*.

Fung C., Tan S., Nakajima M., Skoog E.C., **Camarillo-Guerrero L.F.**, Klein J.A., Lawley T.D., Solnick J.V., Fukami T., Amieva M.R. "High-resolution mapping reveals that microniches in the gastric glands control Helicobacter pylori colonization of the stomach." PLoS biology 17.5 (2019): e3000231.

Contributions

This thesis is the result of my own work except:

- Metagenome assembly, sequence viral prediction with VirFinder and VirSorter, and dereplication at 95% sequence identity was carried out by Alexandre Almeida.
- Read mapping of GPD predictions to 28,060 metagenomes was carried out by Alexandre Almeida.
- Bacterial taxonomic assignment of gut isolates with the GTDB toolkit was carried out by Alexandre Almeida.
- The tool to assign a taxonomic rank to GPD predictions was developed by Guillermo Rangel Pineros.

Abbreviations

Acr	Anti-CRISPR
Abi	Abortive infection
AMG	Auxiliary Metabolic Gene
ANI	Average Nucleotide Identity
ARG	Antibiotic Resistance Gene
BACON	Bacteroidetes-Associated Carbohydrate-binding Often N-terminal
BAM	Bacteriophage Adherence to Mucus
CRISPR	Clustered Regularly Interspaced Short Palindromic Repeats
CTHR	Collagen Triple Helix Repeat
DNA	Deoxyribonucleic acid
FP	False Positive
GPD	Gut Phage Database
GTDB	Genome Taxonomy Database Toolkit
HGT	Horizontal Gene Transfer
HMM	Hidden Markov Model
IBD	Inflammatory Bowel Disease
ICTV	International Committee on Taxonomy of Viruses
ICE	Integrative and Conjugative Element
Ig	Immunoglobulin
ImmeDB	Intestinal Microbiome Mobile Elements Database
KEGG	Kyoto Encyclopaedia of Genes and Genomes
MCL	Markov Cluster
MGE	Mobile Genetic Element
ML	Machine Learning
NCBI	National Centre for Biotechnology information
OMV	Outer Membrane Vesicle
PCA	Principal Component Analysis
PC	Protein Cluster
PD	Parkinson's Disease
PICI	Phage-Inducible Chromosomal Islands
PtW	Piggyback-the-Winner

QC	Quality Control
RBP	Receptor Binding Protein
RNA	Ribonucleic acid
RT	Reverse Transcriptase
RM	Restriction Modification
SaPI	Staphylococcus Aureus Pathogenicity Islands
VC	Viral Cluster
VLP	Viral-Like Particle
VMR	Virus to Microbe Ratio

Table of Contents

Chapter 1: Introduction

1.1 General overview of bacteriophages

1.1.1 The life cycle of bacteriophages

Viruses are the most numerous biological entities on Earth with an estimated population of 10^{31} particles (Brüssow and Hendrix, 2002). Bacteriophages or phages for short, are viruses that infect and replicates within bacteria. Their life cycle beings with the injection of their genome into the cytoplasm of a bacterium followed by either the lytic or lysogenic cycle (Figure 1.1). During the lytic cycle, the cell's metabolism is immediately taken over to replicate the phage DNA and start the synthesis of phage proteins required for the assembly of new viral particles. The cycle finishes when phage lytic enzymes destroy the cell wall and newly formed phages are released from the bacterium (Young, 1992). During lysogeny, the phage genome is either integrated in the bacterium genome or kept as a circular replicon in the bacterial cytoplasm (Lwoff, 1953). At this stage the bacterium is not killed and the carried phage genome is referred to as a prophage while the bacterium becomes a lysogen. Lysogens are able to pass their prophages to daughter cells, however the prophage can be 'awakened' at a future generation and enter the lytic cycle. Phages that exclusively rely on the lytic cycle are called virulent, whereas phages able to enter the lysogenic cycle are called temperate.

Besides the lytic and lysogenic cycles, there are other less studied outcomes of a phage infection. One is displayed by the M13 phage which is able to replicate and generate virions without killing its host (Loh et al., 2019). Another route is when phages are carried inside bacteria but do not integrate or proliferate (pseudolysogeny) (Ripp and Miller, 1997). These phages are inactive in some sense and are asymmetrically segregated upon subsequent divisions. Finally, phages can also accumulate deleterious mutations when integrated in the host genome and as a consequence cannot longer enter the lytic cycle. These defective prophages usually are further degraded, however sometimes a subset of their genes can be beneficial for the host and are conserved (phage domestication) (Bobay et al., 2014).

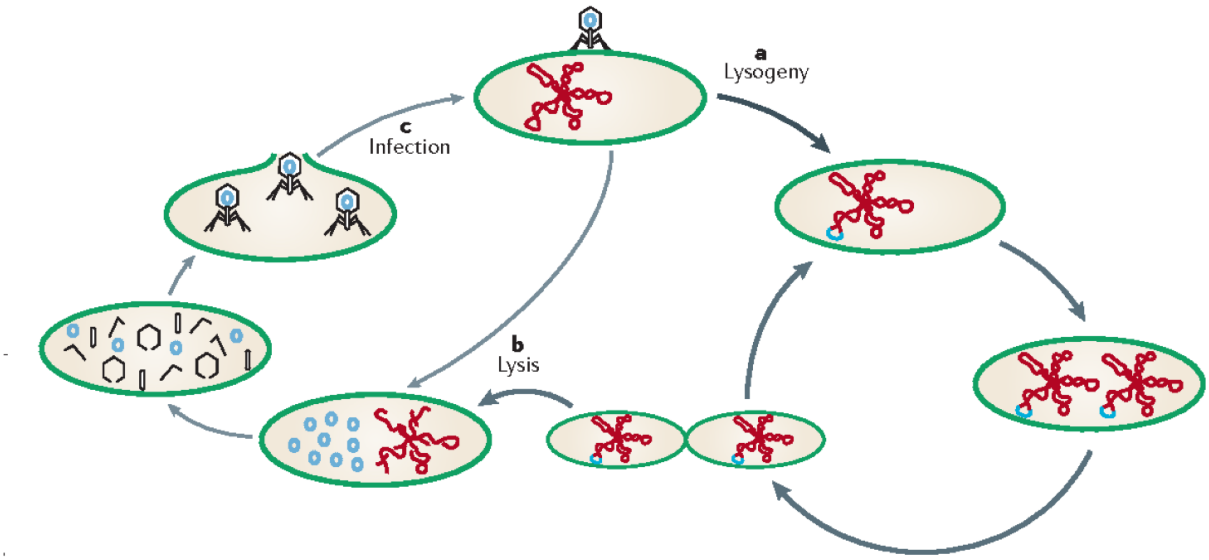


Figure 1.1. The lifecycles of bacteriophages. Lytic and lysogenic are the two main outcomes of a bacterial phage infection. In the former, a phage starts replicating its genome immediately along with the synthesis of phage proteins, ultimately lysing the host and releasing all the newly assembled phages. In the latter, the phage genome integrates into the bacterial genome or is kept as a circular replicon in the bacterial cytoplasm while is passively passed to daughter cells. Sourced from (Reyes et al., 2012)

1.1.2 The outstanding diversity of bacteriophages

Phages can have a DNA or RNA genome (Figure 1.2). However, by far, the most studied phages are those with a linear double stranded DNA (dsDNA) genome. This group is referred to as the *Caudovirales* order and traditionally have been composed of 3 families, namely *Podoviridae, Siphoviridae, and Myoviridae* (Ackermann, 1998). Although with the revised ICTV virus taxonomy from 2019, the *Caudovirales* are now composed of a total of 9 families. A common thing among the *Caudovirales* is the presence of a tail, which is involved in host recognition, cell wall penetration, and genome ejection into the bacteria. *Myoviridae* phages have contractile long straight tails, *Siphoviridae* phages are characterized by non-contractile long flexible tails, and *Podoviridae* phages possess non-contractile short tails. The genomes of *Caudovirales* can vary from 15 kb to 500 kb and are stored in protein complexes called capsids. During virion assembly, 'scaffolding' proteins provide structure for the correct polymerization of the capsid subunits (major capsid proteins) and a connector protein (portal protein) provides a channel for the translocation of the genome into the capsid. Genome packaging is carried out

by a molecular machine composed of the large and small terminases. Replication of DNA generates head-to-tail concatemers of genome units and the small terminase is involved in recognition of phage DNA while the large terminase cuts the DNA concatemer and starts the translocation of DNA fuelled by ATP hydrolysis (Fokine and Rossmann, 2014).

Other less studied phages include the *Tectiviridae* family which possess a linear dsDNA, the *Microviridae* and the *Inoviridae* families which are characterized by having small (<10 kb) and circular single stranded DNA (ssDNA) genomes, the *Leviviridae* family with small (<5 kb) linear ssRNA genomes, and the *Cystoviridae* with dsRNA genomes (Dion et al., 2020).

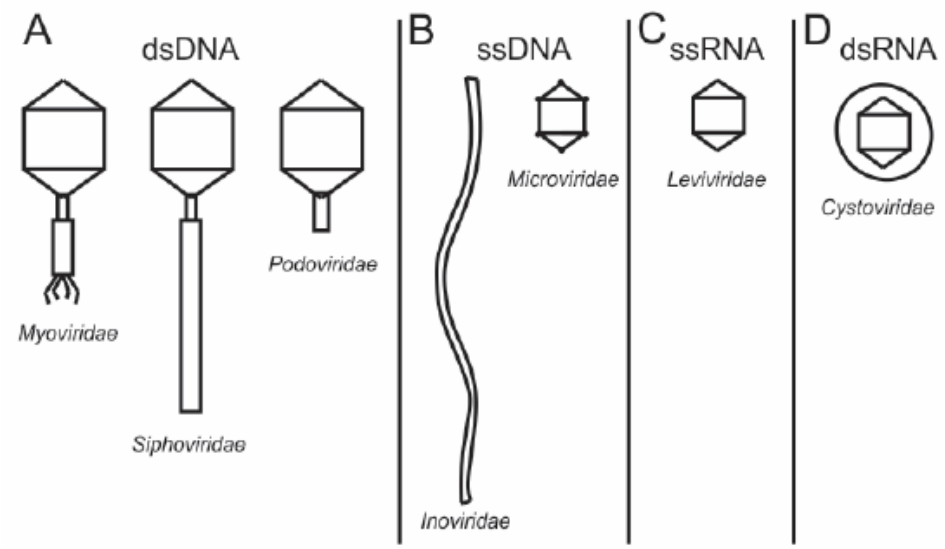


Figure 1.2. The diversity of phages. There is an outstanding phage diversity. Most of the known phages belong to the *Caudovirales* order which traditionally have been divided into 3 families, namely *Myoviridae*, *Siphoviridae*, and *Podoviridae* **(A)**. Other less studied phages include ssDNA phages such as *Inoviridae* and *Microviridae* **(B)**, and phages with an RNA genomes such as *Leviviridae* (ssRNA) **(C)** and *Cystoviridae* (dsRNA) **(D)**. Sourced from (Denton et al., 2013)

1.1.3 Phages: friends or foes of bacteria?

Even though it is tempting to label phages as parasites, they represent a potent force driving ecological functioning and evolutionary change in bacterial communities. A clear example is bacteriophage-mediated horizontal gene transfer, which enhances bacterial adaptive responses to environmental changes (Canchaya et al., 2003). When a prophage undergoes a faulty excision, adjacent chromosomal DNA can end up packaged with the phage genome (specialized transduction) (Morse et al., 1956). A more extreme case can occur when only chromosomal or plasmid DNA is packaged (generalized transduction) (Zinder and Lederberg, 1952).

Phages can also directly increase the fitness of their host. For instance, the viral encoded cI repressor protein which promotes lysogeny of the *E. coli* phage λ, also represses the host gene pckA. This repression in turn causes a decoupling of central metabolism from cellular synthesis, reducing growth rate and may confer a selective advantage in bacteria living in nutrient limited environments (Chen et al., 2005). A more subtle mechanism that phages can use to influence the host phenotype comes from active lysogeny. In this phenomenon, phage excision acts as a regulatory mechanism for expression of bacterial genes without entering the lytic cycle. An example is the phage Φ10403S which its integration disrupts a gene (comK) involved in the escape of its host from the mammalian phagosome. However, when expression of comK is needed, the phage excises and restores the gene function, allowing the survival of its host (Feiner et al., 2015). Other ways bacteria can benefit from phages include the encoding of virulence factors, protection against further phage infection, enhanced biofilm formation, and antibiotic tolerance (Abedon and LeJeune, 2007; Bondy-Denomy et al., 2016; Burmeister et al., 2020; Gödeke et al., 2011).

Co-evolutionary interactions between phages and bacteria also shape the phenotype of bacterial communities. In an effort to prevent successful phage infections, bacteria often mutate and differentially express receptor proteins exploited by phages (Hyman and Abedon, 2010), produce cell surface polysaccharides (Fernandes and São-José, 2018), and can even increase their mutation rate to boost adaptation (Morgan et al., 2010).

1.1.4 The arms-race between phage and bacteria

The Red Queen hypothesis postulates that organisms must constantly evolve and adapt against ever-evolving opposing organisms that share the same environment (Leigh Van Valen, 1973). This scenario is particularly pronounced for bacteria given the constant threat of the lytic cycle and the extremely rapid evolution of phages. Thus, bacteria have developed several strategies to prevent successful phage infections, and at the same time, phages have evolved counter-resistance measures (Figure 1.3).

Bacteria can prevent phage adsorption by altering their receptors (e.g. mutation or chemical modification such as glycosylation) (Harvey et al., 2018) or by masking them with exopolysaccharide capsules (Ohshima et al., 1988). A more indirect approach involves the release of outer membrane vesicles (OMVs) with embedded phage receptors. OMVs thus serve as phage decoys and reduce productive infections (Reyes-Robles et al., 2018). However phages can overcome these hurdles by mutating their receptor-binding proteins (RBPs) to recognize the altered receptors (Meyer et al., 2012), encode multiple RBPs (Schwarzer et al., 2012), or even producing depolymerases to expose a hidden receptor (Fernandes and São-José, 2018).

Even if phages breach extracellular defence mechanisms, bacteria still can counter phages by using intracellular defence systems. Restriction-modification (RM) systems work by cleaving the phage genome upon injection (Oliveira et al., 2014). This is carried out by a restriction endonuclease (R) which recognizes unmethylated phage DNA, while the host DNA remains intact due to methyl modifications by the associated methyltransferase (M). The phage growth limitation (Pgl) system is similar to the RM system except that phages become methylated only after completing the infection cycle (Sumby and Smith, 2002). In a subsequent infection, however, these methylated phages are cleaved upon entry. The DISARM system was recently described and also works by using methylation as an immunity mark, however it provides resistance in the early stages of infection by a yet unknown mechanism (Ofir et al., 2018). Phages have evolved a wide array of strategies to circumvent RM systems (Samson et al., 2013). They can mutate RM sites or modify bases via glycosylation, glucosylation, hydroxymethylation and acetamidation to avoid recognition by the restriction endonuclease. Phages can also activate host methyltransferases or encode their own in order to protect their genome from restriction. Other examples include the Dar system of coliphage P1 which

reduces DNA degradation by interfering with the activity of type I restriction endonucleases and the Ocr protein of coliphage T7 which binds and sequesters the EcoKI endonuclease.

A third type of defence is the CRISPR/cas system which represents a form of adaptive immunity. When a phage infects a bacteria, small fragments of the virus (spacers) are acquired by bacteria. Later on, spacers are transcribed and used as specific probes to recognize phage DNA sequences (protospacers) which leads to degradation by the Cas endonuclease (Barrangou et al., 2007). Phages, on the other hand, can mutate protospacers or modify their bases to avoid recognition by the Cas protein (Paez-Espino et al., 2015), however sometimes escape mutations can lead to phage fitness defects. Anti-CRISPR (Acr) proteins provide a way to overcome this risk by blocking the activity of CRISPR-Cas systems and they do so by mostly interacting with Cas proteins (Bondy-Denomy et al., 2013). As an idea of the complexity of phage/bacteria interactions, CRISPR-cas systems can also be encoded by phages, which can be used to evade host innate immunity (Bondy-Denomy et al., 2013).

In contrast to previous defence systems which focus on protecting individual hosts, abortive infection (Abi) systems act at the population level. They are characterized by allowing phage entry but then the cell host dies in an "altruistic" fashion to severely limit the release of phages and prevent a phage epidemic in the bacterial population (Chopin et al., 2005). Some Abi systems work by exploiting toxin-antitoxin mechanisms. For instance, RnlA is a toxin with endoribonuclease activity which is neutralized by the RnlB antitoxin. Whereas RnlA is a stable protein, RnlB is quickly degraded and thus it needs to be constantly synthesized. However, infection by the T4 phage rapidly shuts off *E. coli* gene expression, resulting in the disappearance of RnlA and allowing RnlB to cause cell death (Naka et al., 2017). Some counter-measures to avoid Abi systems include evolving alternative antitoxins (Otsuka and Yonesaki, 2012), acquiring native antitoxins by recombination with the host (Blower et al., 2012), producing proteins that prevent the degradation of the antitoxin, and directly inhibiting toxins (Alawneh et al., 2016).

Finally, the phage-inducible chromosomal islands (PICIs) are phage parasites that can affect phages by disrupting phage particle assembly and DNA packaging (assembly interference) (Seed, 2015). The best studied members of PICIs are the Staphylococcus aureus pathogenicity islands (SaPIs), which cause mature phage particles to package SaPI DNA rather than phage DNA.

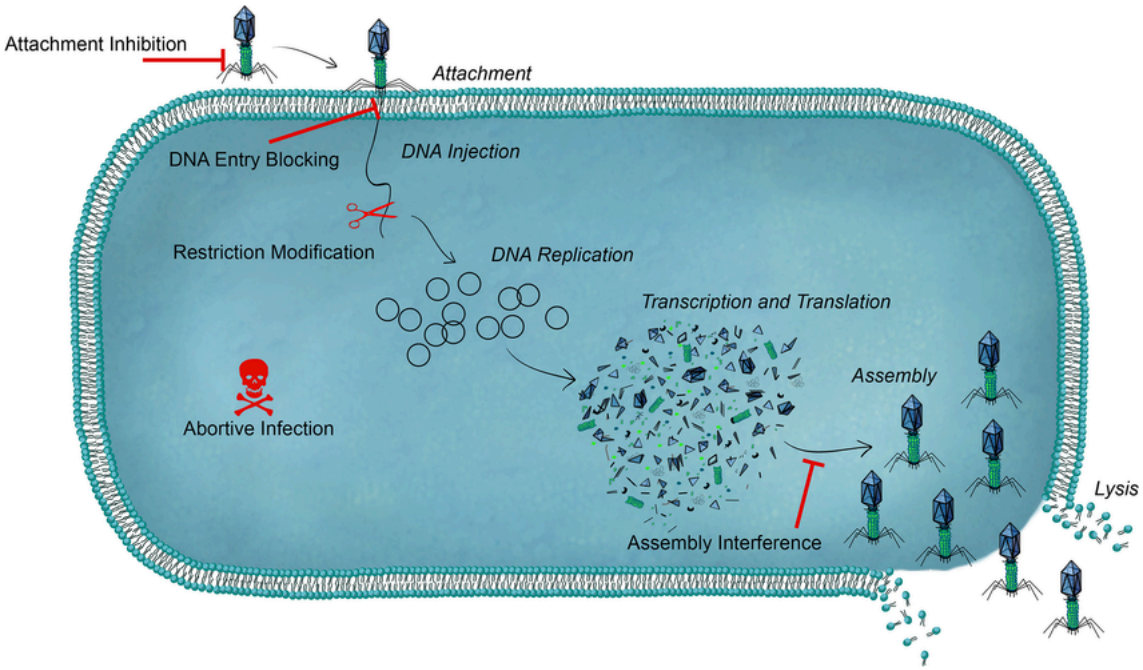


Figure 1.3. Bacterial anti-phage defences. Bacteria have acquired an arsenal of strategies to interfere with phage infections. These defence systems can act by preventing phage attachment and DNA injection, degrading phage DNA by restriction or CRISPR systems, abortive infection, among others. Sourced from (Seed, 2015)

1.1.5 Evolutionary phage-host dynamics

While the previous section highlighted the mechanisms of resistance and counter-resistance, now we review how these strategies vary over time. Two main models have been proposed to explain the dynamics of resistance and counter-resistance. The arms race dynamics model posits that phages select for resistant hosts, which in turn apply selective pressure for phage mutations that restore infectivity, and the cycle repeats. However, coevolutionary experimental studies have shown that the arms race between viruses and bacteria does not continue indefinitely (Hall et al., 2011). One explanation is related to metabolic constraints associated with phage resistance. For instance, if a viral receptor is also a nutrient uptake protein and a resistance conferring mutation impairs nutrient acquisition.

The fluctuating selection model on the other hand, proposes that as the abundance of a fast-growing susceptible host increases, so does the likelihood of encountering a phage, resulting

in increased host mortality and allows for slow-growing resistant bacteria to become majority. However, as the number of phages decreases due to the lack of susceptible hosts, the resistance conferring mutation starts to lose advantage, letting the susceptible fast-growing bacteria to dominate the population and the cycle starts again (Avrani et al., 2012).

1.1.6 Predator-prey dynamics

Whenever we have a predator and a prey interacting, an interesting question arises: how will bacteria and phage populations vary over time? In phage-bacteria interactions two main models have been put forward to explain their dynamics. The first one is the "Kill-the-Winner" model (Thingstad, 2000). This model is based on the assumption that the likelihood of phages killing bacteria is proportional to the relative abundance of the host and mathematically has been approximated with the Lotka-Volterra equations. This way, high levels of bacterial diversity are maintained as overgrown bacteria will be killed by their phages. A second model is "Piggy-back-the-Winner" and it posits that when a host is abundant and growing rapidly, temperate phages will prefer to enter the lysogenic cycle. In addition to replicating "for free" (due to the fast growing rate of its host), they can provide defence against other phages by super infection immunity (Knowles et al., 2016).

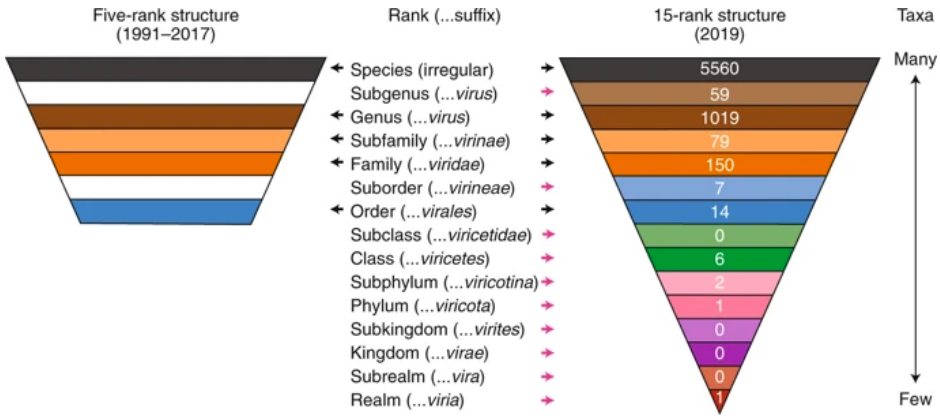
1.1.7 Taxonomy and the recent explosion of phage diversity

The taxonomy of phages is established by the International Committee on the Taxonomy of Viruses (ICTV) which published its first report in 1971. (Adriaenssens and Brister, 2017) Initial classification efforts were based mainly on phage morphology (facilitated by electron microscopy observations) and nucleic acid content, which have been the major criterion for classification at the family taxonomic rank. For many years, most of the phages discovered were categorized to belong to one of the 3 traditional *Caudovirales* families, namely *Podoviridae*, *Siphoviridae*, and *Myoviridae*. However, grouping at lower taxonomic levels such as genus and subfamily was rarely addressed. Demarcation of species in phages is currently set at 95% nucleotide identity, constrained to low levels of genome re-arrangements. In the case of genus, nucleotide identity can drop to 50% as long as the group shares a set of cohesive features such as average genome length, presence of signature genes, average number of tRNAs, etc. Recently, the ICTV has allowed a 15-rank classification which aims to

accommodate the entire spectrum of genetic divergence in the virosphere (Gorbalenya et al., 2020) (Figure 1.4A). This expanded classification matches better the Linnaean taxonomic system. In line with this development, a proposed megataxonomy for all viruses was published this year (Koonin et al., 2020). With this taxonomy current known phages can be placed into other higher orders, for instance, the *Caudovirales* belong to the class *Caudoviricetes*, phylum *Uroviricota*, Kingdom *Heunggongvirae*, and realm *Duplodnaviria*.

With the advent of high-throughput sequencing and metagenomics, there was an explosion on the number of novel phages discovered (Figure 1.4B). With the vast majority of these newly discovered phages only known by sequence, most of them remained unclassified. In an effort to counter this classification issue, several alternative classification schemes were proposed which were based only on sequence information such as the phage proteomic tree, gene-sharing networks, and kmer-based grouping. Proposals to incorporate the vast number of phages discovered by metagenomics into current phage taxonomy are now being considered by the ICTV (Simmonds et al., 2017).

A



B

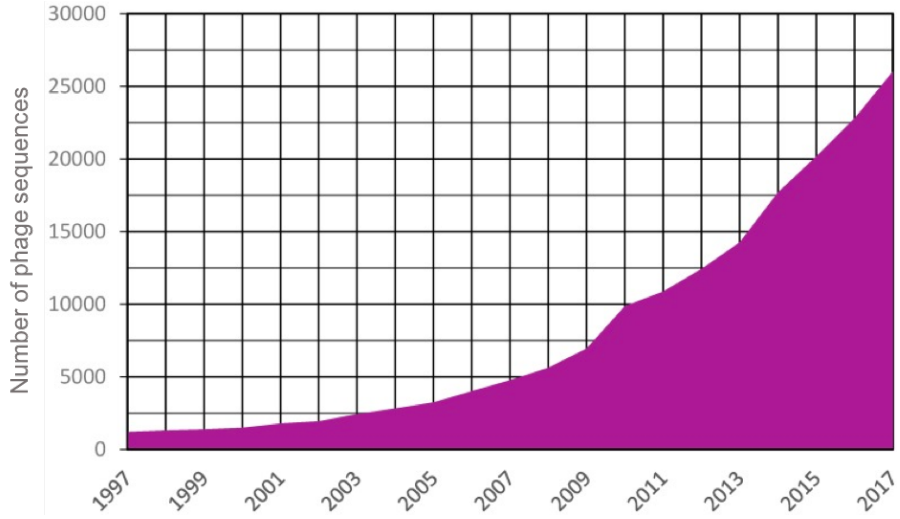


Figure 1.4. Taxonomy of phages. A) The highest taxonomy rank to classify phages was "order". Recently the ICTV incorporated a 15-rank classification which aims to accommodate the entire spectrum of genetic divergence in the virosphere. Sourced from (Gorbalenya et al., 2020). **B)** The number of discovered phage sequences deposited on Genbank across the years was fuelled by high-throughput sequencing and metagenomics. Unfortunately, the majority of sequences remained unclassified. Adapted from (Adriaenssens and Brister, 2017)

1.1.8 Prediction of phages from metagenomic sequences

As mentioned in the previous section, the recent explosion of discovered phage diversity has been fuelled by the mining of metagenomic sequences. A common strategy to identify phages involves the comparison of proteins in the query DNA to a reference database of known phage proteins (Roux et al., 2015). However, this similarity approach is limited to mainly find phages related to the ones in the database, and thus falls short when mining environments with a high level of novel phage diversity. The similarity approach can be improved by the use of Hidden Markov Models, as they are suitable to detect more similarity between novel and known phage proteins.

Another strategy involves the detection of "viral-like" genomic features, such as GC skew, protein length and transcription strand directionality. The use of kmer profiles has also been exploited to differentiate phages from bacterial DNA (Ren et al., 2017).

1.2 Bacteriophages in the human gut

1.2.1 Discovery and isolation of faecal VLPs

Phages in the gut were discovered in 1917 by d'Herelle when he reported "an invisible microbe with antagonistic properties against the Shiga bacillus" in stools from individuals convalescent from bacillary dysentery (D'Herelle, 2007). However, it was not until recently, that more research started to focus on gut phages. In part because of the increased awareness of the gut microbiota in human health, and because gut phages often prey on bacterial hosts which traditionally have been very challenging to cultivate (strict anaerobes) (Browne et al., 2016). Even though now it's technically possible to culture a large number of anaerobic bacteria from the gut, a wealth of information about gut phages has come from the analysis of viral nucleic acids extracted from human faeces. A common procedure, involves the use of 0.2 or 0.45 um filtered faecal samples to greatly reduce non-viral contamination, followed by several physical and enzymatic steps that remove prokaryotic and eukaryotic material (Shkoporov et al., 2018a). The resultant supernatant is enriched in virions, or viral like particles (VLPs) which are then digested to release and sequence the viral nucleic acids. A disadvantage is that VLPs represent only phages that are undergoing the lytic cycle, and thus inactive prophages at the moment of VLP extraction are missed.

1.2.2 Taxonomy of gut phages

Microscopic studies of VLPs and their nucleic acids has shown that the gut phageome is dominated by members of the *Caudovirales* (Hoyles et al., 2014) (Figure 1.5). Other studies have also detected other families such as *Microviridae* and *Inoviridae* (Kim et al., 2011). RNA phages, although present in faeces, are thought to be rare. In addition, giant phages with a genome size > 540 kb in length have been detected in human faeces from Bangladesh. These phages which were assigned a *Prevotella* host, are thought to be enriched in the gut microbiome of individuals who consume non-Western diets (Devoto et al., 2019).

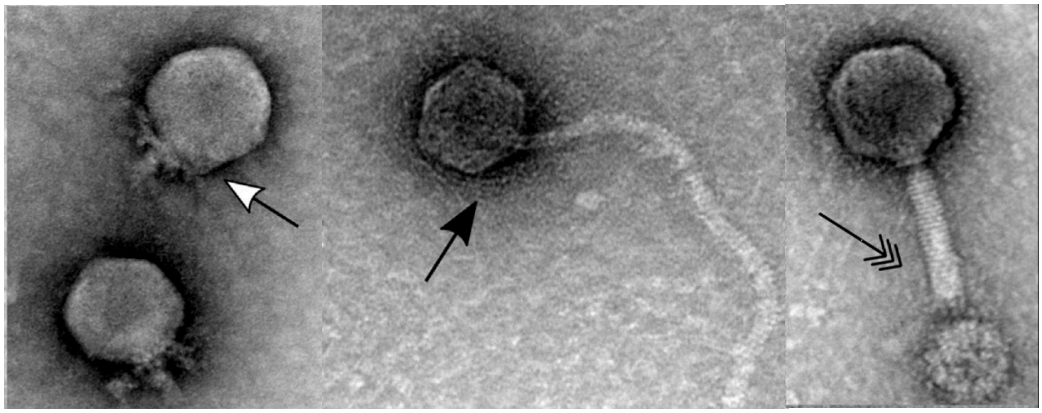


Figure 1.5. Main bacteriophage morphological types detected in a faecal sample. The main phages identified in human faeces belong to the *Caudovirales* order. Here, highlighted from left to right the *Podoviridae*, *Siphoviridae*, and *Myoviridae*. Adapted from (Shkoporov and Hill, 2019)

1.2.3 The case of the crAssphage

The most famous human gut phage is the crAssphage, which was first reported in 2014 and its genome was assembled purely from metagenomic reads (thus the name CRoss-ASSembly) (Dutilh et al., 2014). This phage which is highly prevalent in Western cohorts and can represent up to 90% of the total reads from a single virome, went undetected for years because it represented a completely novel clade of phages. It was later discovered that crAssphage was a member of an expansive bacteriophage family named "crAss-like" which consisted of 4 subfamilies and 10 genera (Guerin et al., 2018). The original member crAssphage belongs to genus I, and it's often referred to as p-crAssphage (prototypical). Its match with CRISPR spacers, the presence of a *Bacteroides* protein domain (BACON) in its genome, and bacterial abundance correlation experiments suggest that p-crAssphage infects a *Bacteroides* species, however its exact host remains elusive to date. On the other hand, a member of genus VI was isolated in the laboratory from *Bacteroides intestinalis* (Shkoporov et al., 2018b)

1.2.4 Phage dynamics in the human gut

It's thought that lysogeny is the predominant lifestyle of phages in the human gut. This is based on the high number of commensal bacteria harbouring prophages (Kim and Bae, 2018), the abundant genes associated with lysogeny in metagenomic studies, the long-term stability of the gut phageome, and low mutation rate over time in temperate-like contigs. (Minot et al., 2013; Reyes et al., 2010a). In addition, some studies have reported relatively low counts of viral particles with 10^9-10^{10} particles per gram of faeces compared to 10^{11}-10^{12} bacteria. Even adjusting for inefficiencies in the purification process, the number of particles still would be in a range of 10^{10}-10^{12} particles per gram of faeces. When taking into account these estimates, the virus to microbe ratio (VMR) in the gut is significantly lower compared to other microbial communities (Manrique et al., 2017).

In addition to the low VMR observed in the gut, the absence of abundance oscillatory patterns of phages and gut bacteria (which are indicative of a kill-the-winner scenario) (Minot et al., 2011), along with the high rate of suggestive lysogeny in the gut, has led to the proposal that Piggyback-the-Winner (PtW) dynamics predominate in the human gut.

However, dynamics between phage and bacteria may deviate from PtW depending on the distance from the intestinal mucus (Figure 1.6). It has been observed that the VMR is in average four times higher in metazoan-associated mucosal surfaces when compared with the surrounding environment (Silveira and Rohwer, 2016). Given that the VMR is positively correlated with the proximity to the intestinal mucus, it has been proposed that lysogeny is favoured at the top mucosal layer, while a lytic lifestyle predominates in the bacteria-sparse intermediary layers (Silveira and Rohwer, 2016). The bacteriophage adherence to mucus (BAM) postulates that metazoan mucosal surfaces and phage co-evolve to maintain phage adherence which limits microbial colonization of the inner layers.

In the case of the infant microbiome, PtW dynamics may not predominate, as there is instability caused by a marked contraction of phage diversity during the first 2 years of life. This type of dynamics aligns better with a kill-the-winner scenario as predicted by the Lotka-Volterra model, which predicts a decay of predators when there is scarce prey (Lim et al., 2015).

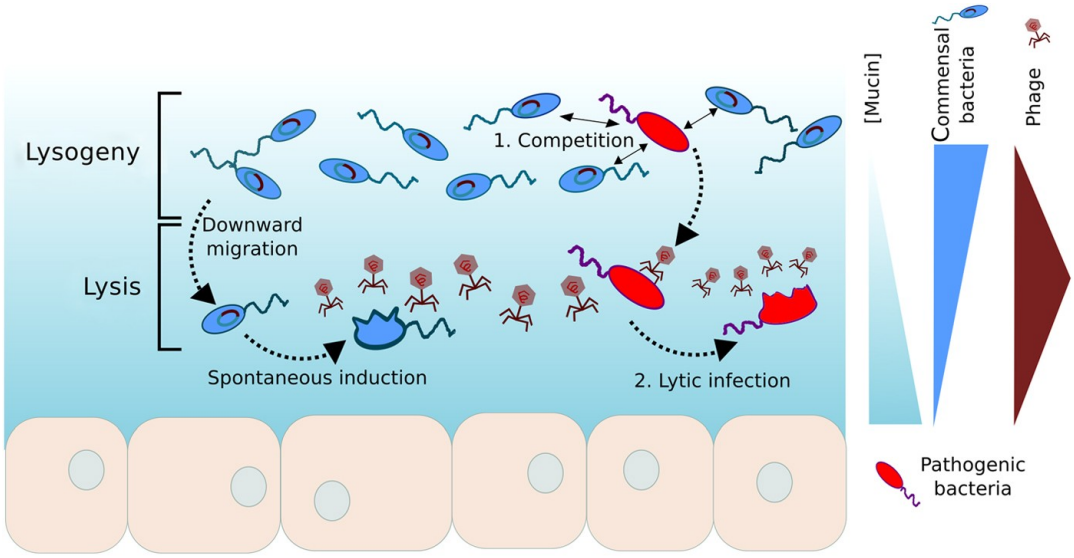


Figure 1.6. Phage dynamics in the human gut. Lysogeny is proposed to be the most prevalent phage cycle in the densely populated human gut (piggyback-the-winner). However, as bacteria move across the intestinal mucus, the lytic cycle is favoured over lysogeny. The bacteriophage adherence to mucus (BAM) postulates that metazoan mucosal surfaces and phage co-evolve to maintain phage adherence which limits microbial colonization of the inner layers. Sourced from (Silveira and Rohwer, 2016)

1.2.5 From lysogeny to the lytic cycle in the gut

Given the high-level of suspected lysogeny in the gut, a key question is whether prophages are active or have become remnants of past phage infections. While there is no a comprehensive study that has evaluated the active fraction of prophages in the gut, a significant proportion of prophages detected by genomic analyses are active (Cornuault et al., 2018; Krupovic and Forterre, 2011; Lugli et al., 2016). In general, phages enter the lytic cycle when they sense a stressor (e.g. activation of the SOS response), it's a survival mechanism that allows them to "abandon a sinking ship". In that regard, induction of gut prophages has been observed by antibiotics (Zhang et al., 2000), diet (such as fructose and short chain fatty acids) (Chatterjee and Duerkop, 2019), bile (Kim et al., 2014), and intestinal inflammation (Diard et al., 2017).

1.2.6 Hosts and host ranges of gut phages

Due to the difficulty of culturing anaerobic gut bacteria, the identity of the hosts targeted by gut phages is a crucial but largely unanswered question. Bioinformatically, CRISPR spacers have been used to link gut phages with predicted hosts. For instance, *Adi et al*. assigned 31 phage contigs to 11 bacterial hosts, with 14 of these phages targeting *Bacteroides* and *Parabacteroide*s (Stern et al., 2012). In another study, one third of 180 phage clusters were linked to abundant taxa such as *Faecalibacteriu*m and *Bacteroides* (Shkoporov et al., 2019). Often phages are restricted to infect single bacterial species, however, intestinal phages may be more promiscuous than expected. For instance, Shkoporov et al. found several phages with broad host range (Shkoporov et al., 2019) and a phage infecting *Faecalibacterium prausnitzii* was shown to also infect *Blautia hansenii* which belongs to a different bacterial taxonomic order (Cornuault et al., 2018). In addition, host range expansion has been observed in a mouse model (De Sordi et al., 2017). However, a study that used a viral tag approach which analysed 363 unique host-phage pairings, found no phages that targeted more than one bacterial species (Džunková et al., 2019). Viral tagging involves the labelling of anonymous virions with a fluorochrome and then they are allowed to attach to host cells. Finally, host-phage pairs are separated by FACS and sequenced to identify the host and the virion. On the other hand, a more comprehensive survey of the host range of gut phages by meta3C proximity ligation (6,651 unique host-phage pairs), found that ~31% of gut phages were not restricted to a single species (Marbouty et al., 2020).

1.2.7 Commonly encoded genes by gut phages

Early insights about the biology of gut phage communities came from the analysis of genetic variation in phage contigs derived from human gut metagenomes (Minot et al., 2012). Hotspots of hypervariation were found in genes homologous to the tail-fibre gene of the Bordetella phage BPP-1, which is hypermutagenized by a unique reverse-transcriptase (RT)-based mechanism (Liu et al., 2002). Moreover, most of the hypervariable loci were linked to genes encoding RTs, highlighting the importance of RTs in the generation of genetic variation for some gut phages.

Other genes that have been found in gut phages are proteins bearing domains from the immunoglobulin (Ig) superfamily. Phages with Ig-like domains have been detected in many

environments, particularly those adjacent to mucosal surfaces. Interestingly, *in-vitro* studies have shown that enrichment of phage in mucus occurs via interactions between Ig-like protein domains and mucin glycoproteins (Barr et al., 2013).

1.2.8 Stability, inter- and intra-diversity of the human gut phageome

The human gut phageome can be defined as the aggregate of phages that inhabit an individual's intestine. It has been found that the human gut phageome is highly diverse between individuals, while intrapersonal variation is minimal and stable (Figure 1.7A,B). In a seminal work (Reyes et al., 2010b), Reyes et al. characterized the faecal viromes of four pairs of adult female monozygotic twins and their mothers by sequencing DNA from VLPs. Analysis of beta diversity revealed that despite remarkable inter-personal variations in their viromes, intra-personal diversity was very low, with >95% of virotypes retained within at least one-year period. Importantly, relative abundances showed minimal variation as well. More evidence about the stability of the gut phageome came from a longitudinal study that monthly tracked the gut phageome of 10 individuals over a period of 1 year by VLP shotgun sequencing. This study revealed that despite certain fluctuations over time, the phageome composition was stable at family and contig level (Shkoporov et al., 2019). This stability was mirrored by the bacterial gut composition which remained stable and individual specific. Another study investigated the relationship between the bacterial microbiome and the virome diversity in 21 adult monozygotic twin pairs (Moreno-Gallego et al., 2019). They found that viromes were unique to individuals, as only 2.83% of the total dereplicated viral contigs were detected in at least 50% of the individuals, and 0.1% were present in all individuals. Notably, this study also showed that phages are the dominant viruses in human gut microbiome, as only 6.42% of the contigs were annotated as Eukaryotic viruses.

The composition of the gut phageome can be altered with diet, however at a lesser degree than interpersonal variation (Minot et al., 2011). Importantly, the variation detected was significantly correlated between bacterial and VLP communities, indicating that diet may affect the gut phageome by perturbing the bacterial gut microbiome.

In contrast to adults, the gut phageome from infants has been found to be less stable. The gut of an infant at birth is considered sterile, but its rapid colonization by microbes derived from

the mother and the surrounding environment leads to the colonization by a phage community. From birth to 2 years of age, there is a contraction and shift in the bacteriophage gut composition, which is in stark contrast with the stable microbiome observed in adults. Moreover, richness and diversity of the gut phageome were found to decrease with age (Lim et al., 2015). Another interesting feature of the infant gut phageome is that the *Caudovirales* and *Microviridae* show an inverse correlation in abundance and diversity up to 2 years of life.

Finally, a controversial concept that has emerged in the field is the existence of a core phageome (Figure 1.7C). Despite the high interpersonal variation found in the gut phageome in previous studies, Manrique et al. proposed that there exists a set of shared phages across individuals referred to as the core phageome (Manrique et al., 2016). In this work, 23 bacteriophages were shared in more than one half of 64 healthy individuals around the world. Moreover, this core set of phages was significantly decreased in individuals with gastrointestinal disease such as IBD. However, a more recent report found that no viral population was detected in more than half of 132 healthy individuals. Specifically, only 1% of phages was shared by over 20% of individuals (Gregory et al., 2019).

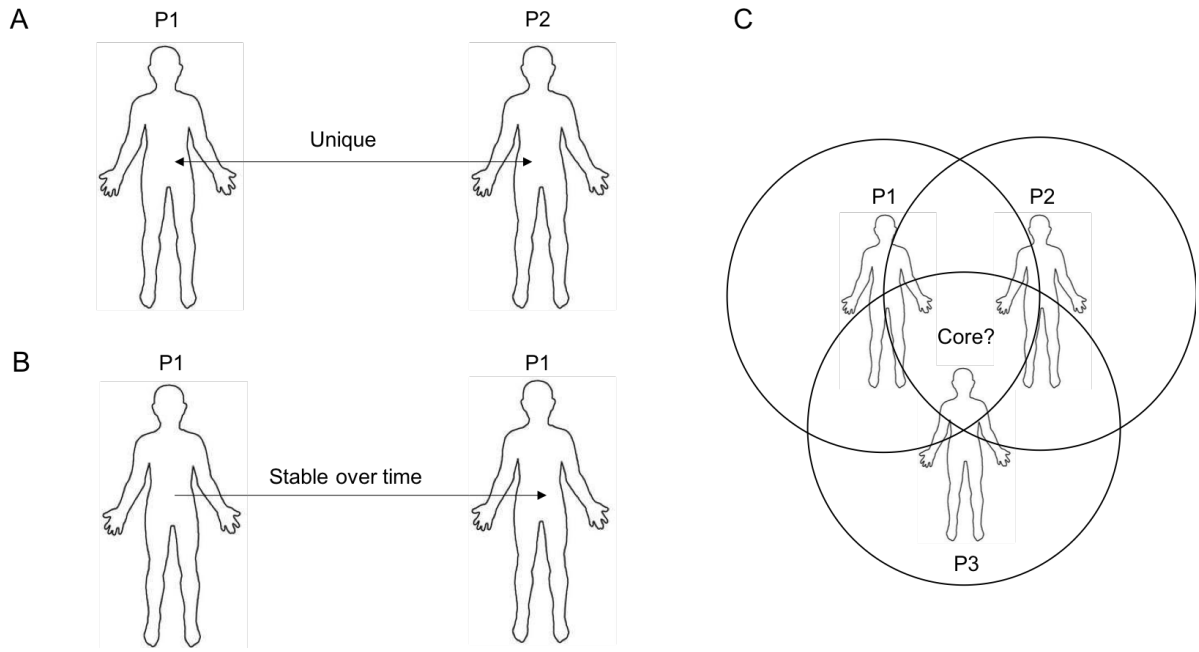


Figure 1.7. Inter- and intra-diversity and stability of the human gut phageome. A) Analysis of phage contigs derived from sequencing faecal viral-like particles (VLPs) has shown that inter-personal variation of the gut phageome is very high among individuals. **B)**

Conversely, the individual gut phageome is stable. **C)** It has been proposed that there is a set of phages shared by a large fraction of individuals, the core phageome. However, this idea is controversial as some studies cannot identify a core phageome.

1.2.9 Gut phages and human disease

Gut phages have been associated with several diseases such as IBD. For instance, it was found that in Crohn's disease and ulcerative colitis the enteric phageome richness increased and that bacterial diversity didn't explain the associated phageome pattern (Norman et al., 2015). However, a subsequent study didn't find evidence of increased phage richness in IBD patients. Instead, it found that healthy controls harboured a stable core of virulent phages that were replaced by temperate phages in Crohn's disease (Clooney et al., 2019).

Another study correlated the increase of strictly lytic virulent lactococcal gut phages with a decrease in Lactococci in Parkinson's disease (PD) patients (Tetz et al., 2018). Lactic acid bacteria are known to produce dopamine and regulate intestinal permeability which are factors implicated in PD pathogenesis. Thus, phages could indirectly contribute to disease by killing beneficial gut bacteria.

Phages can also cause disease by transforming bacteria into pathogens. Certainly, many well-known human diseases are caused by prophage encoded virulence factors such as cholera, diptheria, botulism, and those carrying the Shiga toxin.

1.2.10 Phage therapy

Phages can also be harnessed to treat disease. Shortly after the discovery of phages in 1915, it was realised that they could be used to kill pathogenic bacteria. This idea materialized in 1919 when d'Hérelle first successfully treated several children who were suffering from severe dysentery (Abedon et al., 2011). However, after the discovery of antibiotics, they were disregarded as therapeutic agents particularly in the West (Wittebole et al., 2014). With the rise of antibiotic resistance, there has been a global renewed interest in using phages to treat infections. Unlike antibiotics, phages can be easily mutated to recognize resistant strains, making them very robust to antibiotic resistance; A cocktail of phages can also be used to

mitigate the risk of resistance. In addition, since phages can be very specific to its target strains, there is minimal collateral damage to other bacteria (e.g. gut commensals). Nonetheless, phage therapy also faces some hurdles. For instance, phages can elicit innate and acquired immune responses against them, causing a decrease of their antibiotic activity. The use of temperate phages is inadvisable, given their inherent capacity to the risk of horizontal gene transfer. Phages also contain a large fraction of hypothetical proteins, which could encode proteins that alter bacterial physiology in unexpected ways (Altamirano and Barr, 2019).

Thus, before phages can be deployed as antibiotic agents in different ecosystems such as the human gut, it's necessary to obtain a comprehensive view such as their genomes. Compilation of gut phage genomes could help reveal the function of their genes (e.g. which phages encode virulence factors), identification of the most amenable phages for genetic engineering, their host range, and even assessment of their immunogenicity.

1.3 Thesis aims

The goal of this thesis was to generate critical knowledge about the human gut phageome by harnessing publicly available human gut metagenomes and cultured gut isolates.

Specifically, this thesis aims to:

1) generate the most comprehensive and high-quality database of human gut phage genomes (Chapter 3) ;
2) learn about the functions encoded by gut phages, relevant phage clades, and their bacterial hosts (Chapter 4);
3) investigate global epidemiology patterns of the human gut phageome (Chapter 5).

The objectives relevant to each aim are stated under the introduction of each chapter.

Chapter 2: Methods

2.1 Chapter 3: The Gut Phage Database

2.1.1 Metagenome assembly

Sequencing reads from 28,060 human gut metagenomes were obtained from the European Nucleotide Archive (Leinonen et al., 2011). Paired-end reads were assembled using SPAdes v3.10.0 (Bankevich et al., 2012) with option '--meta', while single-end reads were assembled with MEGAHIT v1.1.3 (Li et al., 2015) both with default parameters.

2.1.2 Viral sequence prediction

To identify viral sequences among human gut metagenomes, VirFinder v1.1 (Ren et al., 2017) which relies on k-mer signatures to discriminate viral from bacterial contigs, and VirSorter v1.0.5 (Roux et al., 2015) which exploits sequence similarity to known phage and other viral-like features such as GC skew were used. While VirFinder is only able to classify whole contigs, VirSorter can also detect prophages and thus classifies viral sequences as 'free' or integrated. Since obtaining high-quality genomes was paramount for downstream analyses, conservative settings for both tools were used. Only metagenome assembled contigs >10 kb in length were analysed for viral prediction. With VirSorter, only predictions classified as category 1, 2, 4 or 5 were considered. In the case of VirFinder, contigs with a score >0.9 and P <0.01 were selected.

Contigs were further quality-filtered to remove host sequences using a blast-based approach. Briefly, the 'blastn' function of BLAST v2.6.0 (Altschul et al., 1990) was used to query each contig against the human genome GRCh38 using the following parameters: '-word_size 28 -best_hit_overhang 0.1 -best_hit_score_edge 0.1 -dust yes -evalue 0.0001 -min_raw_gapped_score 100 -penalty -5 -perc_identity 90 -soft_masking true'. Contigs with positive hits across >60% total length were excluded.

2.1.3 Sequence clustering

Dereplication of the filtered contigs was performed with CD-HIT v4.7 (Li and Godzik, 2006) using a global identity threshold of 99% ('-c 0.99'). This was performed first on contigs obtained within the same ENA study, and afterwards among those obtained across studies. A final set of representative viral sequences was generated by clustering these resulting contigs at a 95% nucleotide identity over a local alignment of 75% of the shortest sequence (options '-c 0.95 -G 0 -aS 0.75').

2.1.4 Quality control of GPD predictions

In order to ensure a high-quality of GPD predictions I removed integrative and conjugative elements by using a machine learning approach.

The training set consisted of all experimental ICEs with intact sequence retrieved from ICEberg 2.0 (Bi et al., 2012) and the phage RefSeq genomes from NCBI (Brister et al., 2015). The test set was downloaded from the Intestinal microbiome mobile elements database (ImmeDB) (Jiang et al., 2019) corresponding to the "ICEs" and "Prophages" datasets. By parsing GFF files with custom Python scripts, for each sequence I calculated 3 high-level features, namely number of genes/kb, number of hypothetical proteins/total genes, and 5-kmer relative frequency (4^5 = 1024 kmers). I used Keras with the TensorFlow (Abadi et al., 2016) backend to train a feedforward neural network with an initial hidden layer of size 10 (ReLU activation), followed by another hidden layer of size 5 (ReLU activation) and a final neuron with a sigmoid activation function. Model selection was carried out with 5-fold cross-validation. I trained the network using the Adam optimizer and the binary cross entropy as the loss function.

I carried out the classification by allowing a false positive rate of 0.25% with a recall of 91%. Finally, I excluded genomes that were predicted to belong to non-phage taxa (82 predictions)

The code for the classifier can be found here:
https://github.com/cai91/GPD

2.1.5 Genome completeness and contamination

Genome completeness and contamination was evaluated by running CheckV v0.5.1 (Nayfach et al., 2020) with the "end_to_end" program.

2.1.6 Viral taxonomic assignment

Viral taxonomic assignment of contigs was performed using a custom database of phylogenetically informative profile HMMs (ViPhOG v1, available here: ftp://ftp.ebi.ac.uk/pub/databases/metagenomics/viral-pipeline/hmmer_databases), where each model is specific to one viral taxon. First, protein-coding sequences of each viral contig were predicted with Prodigal v2.6.3 (Hyatt et al., 2010). Thereafter, 'hmmscan' from HMMER v3.1b2 (Eddy, 1998) was used to query each protein sequence against the ViPhOG database, setting a full-sequence E-value reporting threshold of 10^{-3} and a per-domain independent E-value threshold of 0.01. Resulting hits were analysed to predict the most likely and specific taxon for the whole contig based on the following criteria: (i) a minimum of 20% of genes with hits against the ViPhOG database, or at least two genes if the contig had less than 10 total genes; and (ii) among those with hits against the ViPhOG database, a minimum of 60% assigned to the same viral taxon.

2.1.7 Clustering of phages into VCs

I first created a BLAST database (makeblastdb with options -parse_seqids -dbtype nucl) of all the nucleotide sequences stored in GPD and then carried out all the pairwise comparisons by blasting GPD against itself (I only kept hits with evalue<=0.001). Then, for every pairwise comparison, I calculated the coverage by merging the aligned fraction length of the smaller sequence that shared at least 90% sequence similarity. I kept only the results with a coverage >75%. Finally, I carried out a graph-based clustering by running the Markov Clustering Algorithm (MCL) (Dongen (S.M.), 2000) with an inflation value of 6.0

2.1.8 Bioinformatics tools

The code for the tools developed in this work can be found here:

DotBlast: https://github.com/cai91/dotBlast
HyperVir: https://github.com/cai91/hyperVir
vMatch: https://github.com/cai91/vMatch

2.2. Chapter 4: Function, phylogeny and host assignment of gut phages

2.2.1 Detection of function in gut phages

KEGG pathways, modules, and orthologs were predicted with eggNOG-mapper V2.0.0 (Huerta-Cepas et al., 2017) . Annotation of predictions was carried out using Prokka v. 1.5-135 (Seemann, 2014).

2.2.2 Clustering of proteins into protein clusters (PCs)

I predicted the whole proteome of GPD with Prodigal v2.6.3 (metagenomic mode) (Hyatt et al., 2010) and masked the low-complexity regions with DustMasker. I then created a BLAST (Altschul et al., 1990) database of all the protein sequences and carried out all the pairwise comparisons by blasting the GPD proteome against itself (E-value<=0.001). Then, for every pairwise comparison, I calculated a similarity metric as defined by Chan et al (Chan et al., 2013). Finally, I ran the Markov Clustering Algorithm (MCL) (van Dongen, 2000) with an inflation value of 6.0 and removed clusters with only 1 member.

2.2.3 Phylogenetic analyses

The phylogenetic tree comparing Gubaphage against crAss-like phages was constructed by aligning the corresponding large terminase genes with MAFFT v7.453 (Katoh et al., 2002) –auto mode, followed by FastTree v2.1.10 (Price et al., 2010). The results tree was visualized on iTOL (Letunic and Bork, 2007). For studying the phylogenetic structure of Gubaphage and *Picovirinae*, I calculated the fraction of shared protein clusters among all the Gubaphage genomes and then carried out hierarchical clustering with average linkage and Euclidean metric.

2.2.4 Taxonomic assignment of bacterial genomes

Bacterial isolate genomes were taxonomically classified with the Genome Taxonomy Database Toolkit (GTDB-Tk) v0.3.1 (Chaumeil et al., 2019) (https://github.com/Ecogenomics/GTDBTk) (database release 04-RS89) using the

'classify_wf' function and default parameters. Taxa with an alphabetic suffix represent lineages that are polyphyletic or were subdivided due to taxonomic rank normalization according to the GTDB reference tree. The unsuffixed lineage contains the type strain whereas all other lineages are given alphabetic suffixes, suggesting that their labelling should be revised in due course.

2.2.5 Host assignment

I predicted CRISPR spacer sequences from the 2898 gut bacteria using CrisprCasFinder-2.0.2 (Couvin et al., 2018). I only used spacers found in CRISPR arrays having evidence levels 3 and 4. I assigned a host to a prediction only if the putative host CRISPR spacer matched perfectly to the phage prediction (100% sequence identity across whole length of CRISPR spacer). I carried out the screen by blasting all the predicted CRISPR spacers against the nucleotide GPD BLAST database using the following custom settings (task: blastn-short, -gapopen 10, -gapextend 2, penalty -1, -word_size 7m -perc_identity 100). I retained only hits that matched across the whole length of the spacer with a custom script. In addition, prophages were assigned to the bacterial assembly from which they were predicted. In order to assess the prevalence of false positives due to random chance, I generated 100 sets of CRISRPR random spacers and mapped them against the GPD.

2.2.6 Assessing viral diversity patterns

To compare viral diversity patterns across different gut bacteria, the number of VCs that targeted each bacterial genus was normalized by the total number of isolates from that genus. A VC was considered to target a gut isolate if at least 1 of the genomes from the cluster was predicted to infect it by either CRISPR matching or prophage assignment.

2.2.7 Host range analysis

The number of VCs restricted to target a bacterial taxonomic rank (e.g. species, genus, family) was calculated by predicting all the bacterial hosts associated to each VC and then computing the set for each rank. If the set was a singleton, then the VC was considered to be restricted to that bacterial taxonomic rank.

The gut bacteria isolate tree showing broad host range VCs was constructed by considering all the VCs not restricted to a single genus (cross-family). For each VC, a pair of bacteria assemblies that matched the different genera were picked. The tree was visualized on iTOL.

2.3 Chapter 5: Global distribution and epidemiology of gut phages

2.3.1. Metagenomic read mapping

To estimate the prevalence of each viral species, metagenomic reads were mapped using BWA-MEM v0.7.16a-r1181 ('bwa mem -M') (Li and Durbin, 2009) against the GPD database (clustered at 95% nucleotide identity). Mapped reads were filtered with samtools v1.5 (Li et al., 2009) to remove secondary alignments ('samtools view -F 256') and each viral species was considered present in a sample if the mapped reads covered >75% of the genome length.

2.3.2 Correlation of phages detected and sample sequencing depth

The number of phages detected was calculated by counting the number of GPD genomes that mapped to each of the 28,060 metagenomic samples and then associating it with the corresponding sample sequencing depth. Pearson's r was calculated with the function *stats.personr* from the Python package SciPy v1.3.1

2.3.3. Geographical distribution of metagenomic samples

Similarity between 2 samples was calculated by computing the number of shared VCs divided by the total number of VCs in both samples (Jaccard index). Only deeply sequenced samples (>50 million reads) and healthy samples were considered for the analysis. Distribution of samples was visualized with principal component analysis (PCA) using the *decomposition.PCA* function from scikit-learn v0.22.2. Confidence ellipses encompass 2 standard deviations for each lifestyle samples. PERMANOVA test was carried out with *stats.distance.permanova* function from the Python library scikit-bio v0.5.6

2.3.4 Calculation of phage carriage

Phage carriage was calculated by counting the number of different VCs found in each of the deeply sequenced samples (>50 million reads) for each continent. The Mann Whitney U-test was used to test significance with the *stats.mannwhitneyu* function from the Python package SciPy v. v1.3.1

2.3.5 Detection of enterotypes targeted by VCs

For each analysed region (North America, South America, Europe, Africa, Asia, Fiji and Australia), I predicted all the aggregate bacterial genera targeted by the corresponding genomes that mapped to each region. I then counted the number of genomes that targeted *Bacteroides* genera (*Bacteroides*, *Bacteroides* A, *Bacteroides* B) or the Prevotellaceae family (*Prevotella*, *Paraprevotella*) and normalized by total targeted genera found in each region. Statistical testing was carried out with the *stats.chisquare* function from SciPy v1.3.1.

2.3.6 Network of globally distributed phages

Globally distributed phages were detected by screening VCs for which at least 1 genome of the cluster was found in at least 5 continents. The host-phage network was generated by drawing an edge between each global VC and the predicted bacterial genera they infected. The network was visualized with Cytoscape v3.6.1.

2.3.7 Core virome analyses

In order to evaluate how many VCs covered a specific proportion of samples, I calculated how many samples contained at least 1 VC from a set of VCs. A VC was considered to be found in a sample if at least 1 of the genomes of a VC mapped to the sample. I repeated this procedure with sets sizes ranging from 1 to 500 VCs. Sets grew following the rank of the VCs from biggest to lowest (by number of genomes). When considering the crAss-like family, Gubaphage, and *Picovirinae* clades, I considered them present in a sample if any of the genomes associated to these clades mapped to the sample.

2.4 GPD resource and metadata

GPD genomes and associated metadata can be found here:
http://ftp.ebi.ac.uk/pub/databases/metagenomics/genome_sets/gut_phage_database/

Chapter 3: The Gut Phage Database

3.1 Introduction and aims

The first metagenomic studies revealed that the majority of the viral gut diversity is novel (81%-93%) (Manrique et al., 2016; Reyes et al., 2010), and since only recently their bacterial hosts started to be cultured (Browne et al., 2016), gut phage host assignment and host range have remained largely uncharacterized. An exception has been crAssphage, a phage discovered in 2014 by computational analysis of metagenomic reads and found in >50% of Western human gut microbiomes (Dutilh et al., 2014). A surprising finding was that the majority of phage sequences uncovered by metagenomics could not be classified into any known viral taxonomy laid out by the International Committee on Taxonomy of Viruses (ICTV) (e.g. species, genus, family), prompting many researchers to organize phage predictions from metagenomic datasets into custom grouping schemes based solely on genomic features (Bin Jang et al., 2019).

More recently, gut metagenomes have been mined in order to compile a more comprehensive list of gut phage genomes (Gregory et al., 2019; Paez-Espino et al., 2019). Nevertheless, the limited number (<700) of metagenomes used to construct these databases, and the median fragment size of their predictions (<15 kb as opposed to ~50 kb for an average *Caudovirales* phage genome), suggests that we have yet to capture a globally representative gut phage diversity and the current phage genomes are likely far from complete. Indeed, a recent report estimated that the IMG/VR database, which contains viral sequences from a wide range of environments including the human gut, showed that only 1.9% of the predictions were complete, and 2.5% high-quality (Nayfach et al., 2020). These issues highlight the need for a comprehensive resource of longer and complete reference phage genomes to enable genome-resolved metagenomics for virome studies.

In this chapter, I describe the construction of the largest database to date that harbours the human gut phage sequences, which were product of mining 28,060 metagenomes and 2898 isolate genomes derived from the human gut microbiota. I investigate ways to organise the huge viral diversity uncovered in this work in order to improve the characterisation of gut

phages in the following chapters. I also developed tools that can aid in the exploratory analysis of viral genomes that will be presented in this chapter.

The aims of the research presented in this chapter are:

- generate the Gut Phage Database (GPD), a high-quality and comprehensive database of the human gut bacteriophage sequences;
- group viral diversity into meaningful clusters to enable more powerful downstream analyses;
- Develop tools for the high-throughput analysis of genome synteny, hypervariation, and phylogeny of viral genomes.

3.2 Results and discussion

3.2.1 Construction of the gut phageome database (GPD)

In order to uncover the diversity of human gut bacteriophages, the biggest datasets of human gut metagenomes (n=28,060) and reference genomes of cultured gut bacteria (n=2,898) were mined. In addition, the metagenomes had a worldwide distribution, as they originated from 28 different countries spanning six major continents (Africa, Asia, Europe, North America, South America and Oceania). To identify viral sequences among human gut metagenomes, over 45 million contigs were assembled and screened with VirFinder (Ren et al., 2017), which relies on *k*-mer signatures to discriminate viral from bacterial contigs, and VirSorter (Roux et al., 2015), which exploits sequence similarity to known phage and other viral-like features such as GC skew. Since obtaining high-quality genomes was paramount for downstream analyses, conservative settings were used for both tools and only predictions that were at least 10 kb long were kept. After removing contamination with a machine learning approach (see below) and dereplicating the final set of filtered sequences at a 95% nucleotide identity threshold (over a 75% aligned fraction), a database of 142,809 gut phage sequences was generated (the gut phage database, hereafter referred to as GPD) (Figure 3.1).

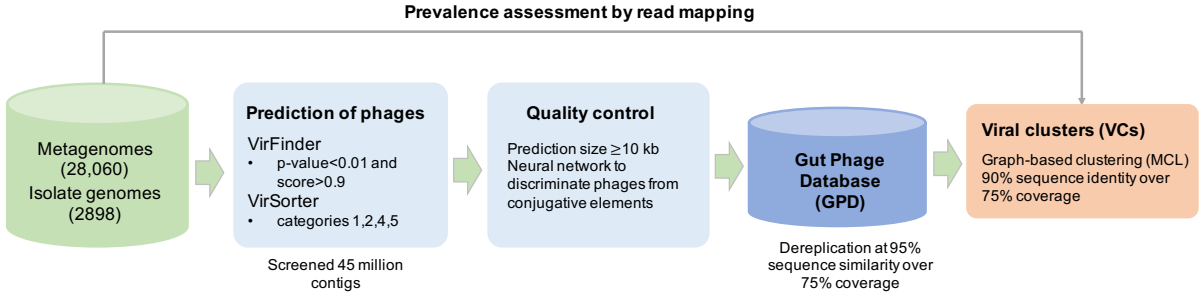


Figure 3.1. Generation of the Gut Phage Database (GPD). An initial dataset composed of 28,060 public human gut metagenomes and 2898 gut bacteria isolate genomes were mined to identify phage genomes. After assembling 45 million contigs, predictions were carried out with VirFinder and VirSorter. Whereas the former is only able to process whole contigs, the latter can also detect integrated viral sequences or prophages. In order to minimize false positives, conservative setting were used for both tools and only fragments > 10 kb were kept. A neural network was trained to remove further contamination caused by ICEs. Predictions were dereplicated at 95% nucleotide identity and they were stored in the gut phage database. In order

to further organize viral diversity, predictions were grouped into viral clusters (VCs). Finally, read mapping was used to quantify prevalence of VCs in the original metagenomes (epidemiology results in Chapter 5).

3.2.2 Decontamination using a machine learning approach

Many false positives (FPs) gene predictions coded for type IV secretion systems and relaxases, suggesting contamination by conjugative mobile elements (Guglielmini et al., 2013). Although plasmids can encode Type IV machinery, I decided to focus on integrative and conjugative elements (ICEs) as conjugation is an inherent feature of their lifestyle (Delavat et al., 2017). In a sense, ICEs behave like temperate "intracellular phages": they integrate into a bacterial genome, can excise from the chromosome and encode a tail-like structural machinery necessary for injecting their DNA into another host. Thus, it's understandable that some of them can be predicted as phages. However, given the widespread use of VirFinder and VirSorter, it came as a surprise that previous reports that used these tools never discussed or raised a warning about potential contamination by conjugative elements. This contamination issue was further exacerbated because many predictions contained truncated ICEs and uncharted diversity, making difficult to discriminate by a marker gene approach.

In order to automate the detection of FPs, I devised a machine learning approach to carry out a further round of decontamination. A feedforward neural network was used to discriminate phages from ICEs. Gene density (genes/kb), kmer signature (pentanucleotide composition), and fraction of hypothetical proteins (hypothetical genes/total genes) were selected as machine learning features, since these metrics can be computed for incomplete sequences and do not rely on direct specific homology (Figure 3.2A and 3.2B). In general, phages had higher densities of genes and hypothetical proteins. The former could be attributed to a selective pressure of phages of fitting their genome into the capsid, while the latter could be explained by poor annotation of phage structural proteins due to their lack of conservation (Seguritan et al., 2012). The extent of discrimination of phages from ICEs by computing these two metrics can be appreciated in Figure 3.2C where they clearly segregate (phages in blue and ICEs in red). The classifier was trained with validated experimental sequences of phages (RefSeq, n=2,387) and ICEs (ICEberg 2.0, n=113). Model selection was carried out with 5-fold cross-validation and the classifier showed an excellent performance in an independent test set

(AUC>0.97) harbouring human gut mobile genetic elements (MGEs) (Figure 3.2D). I carried out the classification by allowing a false positive rate of 0.25% with a recall of 91%.

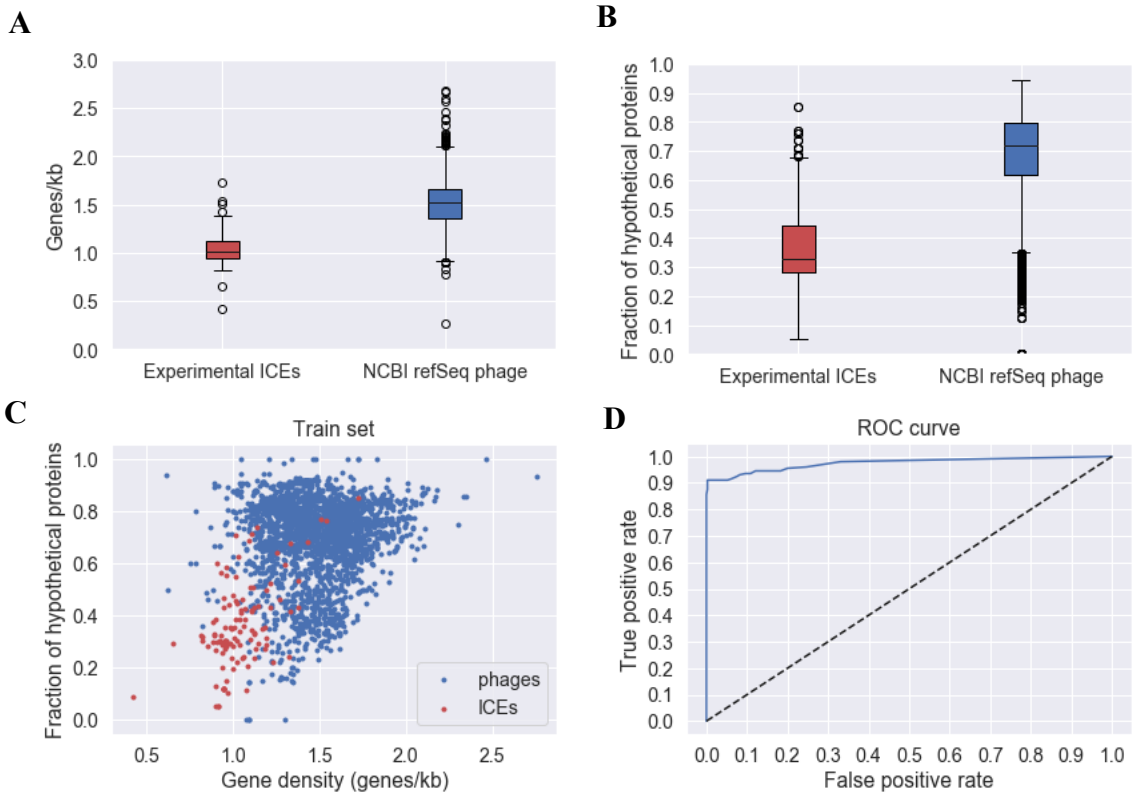


Figure 3.2 – A machine learning approach to distinguish phages from ICEs. In order to discriminate ICEs from phages I relied on three features: kmer signature, gene density, and fraction of hypothetical proteins. Kmer signature has already been exploited as a way to discriminate phages from host DNA. Generally, gene density **A)** and fraction of hypothetical proteins **B)** were lower for ICEs than for phages. **C)** When experimental sequences of ICEs (in red, n =113) and genomes of NCBI phages (in blue, n=2,387) are described by these two features, they clearly segregate. I trained a feed forward neural network that harnessed the 3 features described using experimental sequences from ICEs and phages and benchmarked it with a dataset of gut phages (n=201) and ICEs (n=405). **D)** The classifier had an excellent performance in an independent dataset with an AUC>0.97.

3.2.3 GPD significantly expands gut bacteriophage diversity

In order to assess the viral diversity of the GPD at high taxonomic levels, I used a graph-based clustering approach to group genetically related phages. Merging GPD with RefSeq and two other human gut phage databases (GVD and IMG/VR) (Gregory et al., 2019; Paez-Espino et al., 2019), resulted in the generation of 21,012 non-singleton viral clusters (VCs) with at least 1 GPD prediction (GPD VCs). A VC corresponds to a viral population sharing approximately 90% sequence identity over ~75% aligned fraction.

Comparison of GPD against RefSeq phage genomes, revealed only 171 out 21,012 VCs overlaps. Phages from these 171 VCs mainly infect *Escherichia*, *Enterobacter*, *Staphylococcus*, and *Klebsiella* genera, reflecting the bias of the RefSeq database to harbour phages from well-known clinically important and traditionally culturable bacteria. Consistent with previous reports of phage predictions from metagenomic datasets (Hoyles et al., 2014), I was not able to confidently assign a family to the majority (~80%) of GPD VCs, while the rest corresponded mainly to the *Podoviridae*, *Siphoviridae* and *Myoviridae* families (Figure 3.3A). These 3 viral families belong to the *Caudovirales* order (phages characterized by having tails and icosahedral capsids) which from microscopic studies have been found to be enriched in human faeces (Hoyles et al., 2014; Roux et al., 2012).

For comparison purposes, in addition to GPD VCs, I also considered VCs without GPD predictions (Figure 3.3B). Analysis of VCs composed from only GPD and IMG/VR genomes showed 3,699 overlaps, while I found 3,206 VCs composed of only GPD and GVD genomes. Moreover, GPD harboured the highest number of unique VCs with 12,731 novel clusters. On the other hand, 1099 VCs, and 113 VCs were unique to IMG/VR and GVD, respectively. In addition, 1205 VCs were shared by the three databases. Interestingly, the number of VCs with an assigned phage taxon was lower in the VCs that were unique to GPD as opposed to those shared with GVD and IMG/VR (18.74% vs 27.8%) (P = 1.96e-9, χ^2). Thus, GPD considerably increases the known gut phage diversity in the human gut. This phage diversity expansion is likely driven by the high number of gut metagenomes mined and their global distribution which allows the retrieval of rarer gut phage clades.

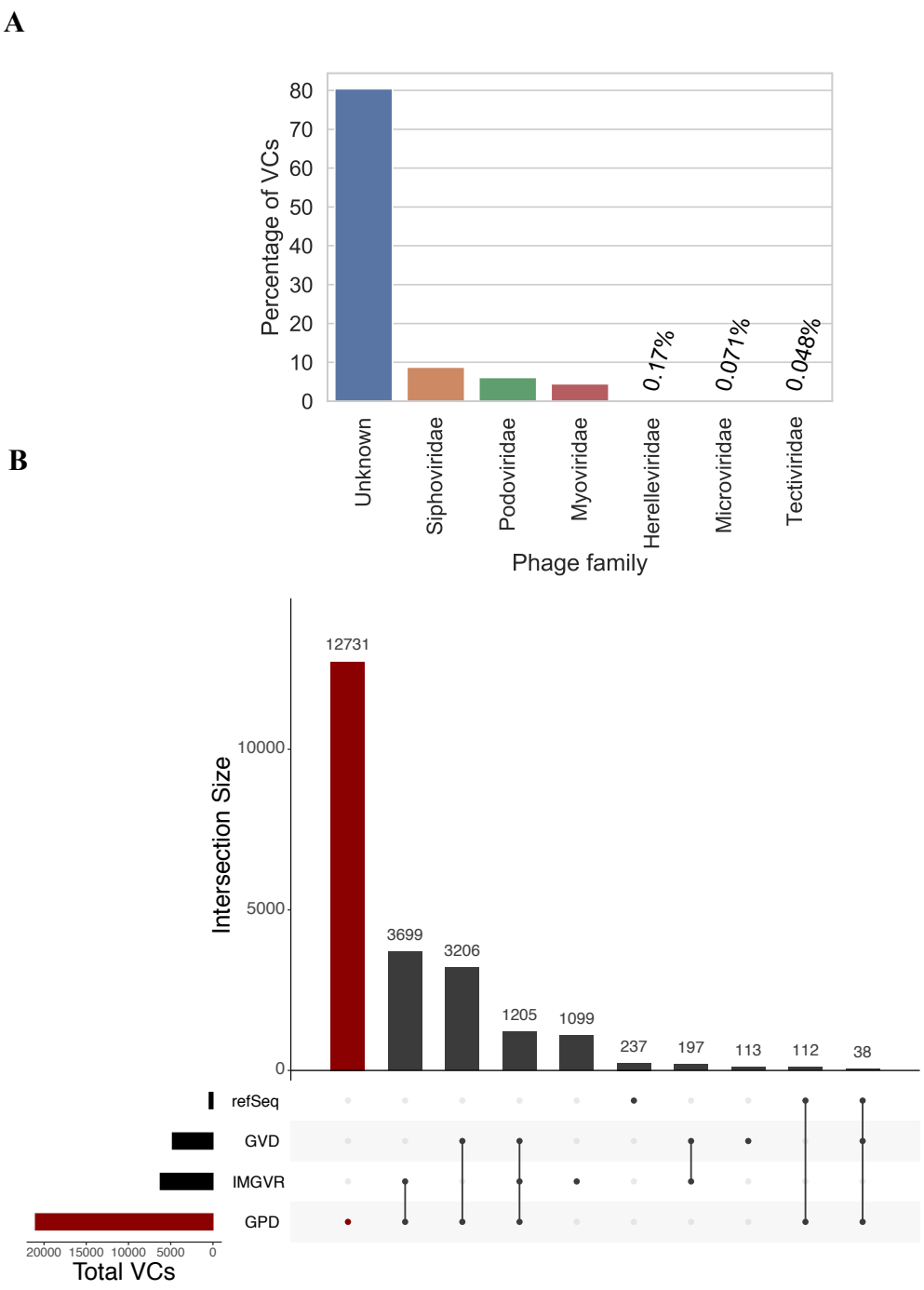


Figure 3.3. GPD taxonomy assignment and comparison against other gut phage databases. **A)** Most of GPD VCs (~80%) could not be assigned to a phage family. The assigned fraction corresponded to mainly families of the *Caudovirales*. **B)** UpSet plot comparing GPD against other public gut phage databases. GPD captures the greatest unique diversity of phage genomes that inhabit the human gut.

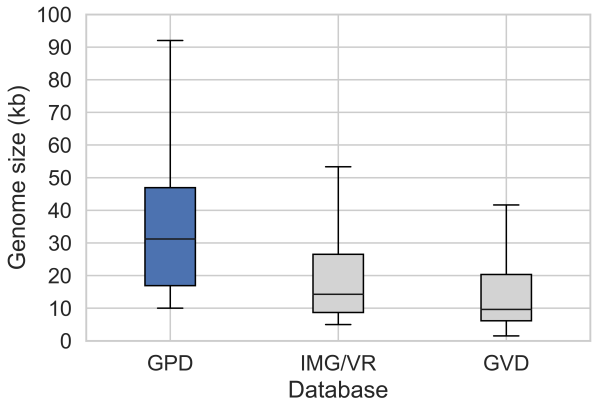
3.2.4 Genome completeness of GPD

Genome completeness is another important feature of a high-quality reference genome database. Unlike prokaryotic genomes, there is no current consensus tool to assess phage completeness and contamination, thus multiple complementary approaches were explored to assess the GPD genome completeness. First, I assessed genome size. The *Caudovirales* order, which is considered a dominant group of the human gut phageome, possesses an average genome size of ~50 kb (Ackermann, 1998). Based on this criteria, GPD harbours the most complete gut phage genomes as it has the largest median genome size with ~31 kb, followed by IMG/VR and GVD with 15 and 11 respectively (Figure 3.4A).

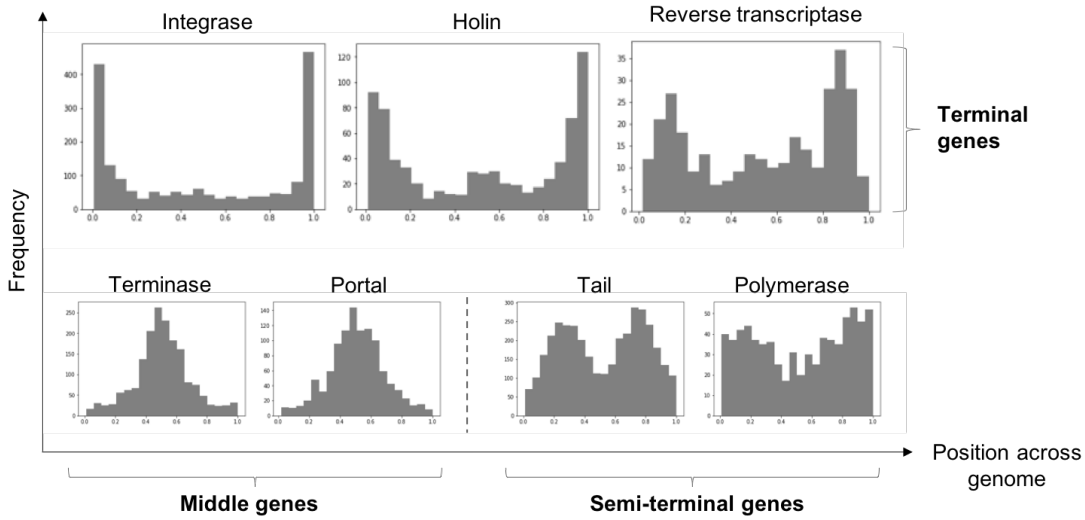
I further assessed completeness by studying the genome organisation of the GPD phage. Figure 3.4B shows the consensus position of marker genes along GPD genomes. I found that key marker genes localized at their expected positions within the predictions. For instance, integrases were more often found at the edges (terminal genes), terminases in the middle, and polymerases in between (semi-terminal genes). This observation reflects the highly complete nature of the GPD genomes. Moreover, this result highlighted the large number of linear genomes which can be a result of prophages or an inherent feature of a phage clade (e.g. *Caudovirales*)

Finally, I estimated the level of completeness of each viral genome using CheckV (Nayfach et al., 2020) (Figure 3.4C). This tool estimates the expected genome length of a viral prediction based on the average amino acid identity to a database of complete viral genomes from NCBI and environmental samples. In total, 41,248 (29%) of the viral genomes were classified as high quality (of which 13,249 were predicted to represent complete genomes), 38,574 (27.01%) as medium quality, 53,116 (37.19%) as low quality, and 9,691 (6.78%) as non-determined. The median genome completeness of all genomes stored in the GPD was estimated to be 63.5% (interquartile range, IQR= 34.68%–95.31%) (Figure 3.4D). Estimation of non-viral DNA by checkV showed that 73.5% of GPD predictions had no contamination whereas 84.13% had a predicted contamination <10%.

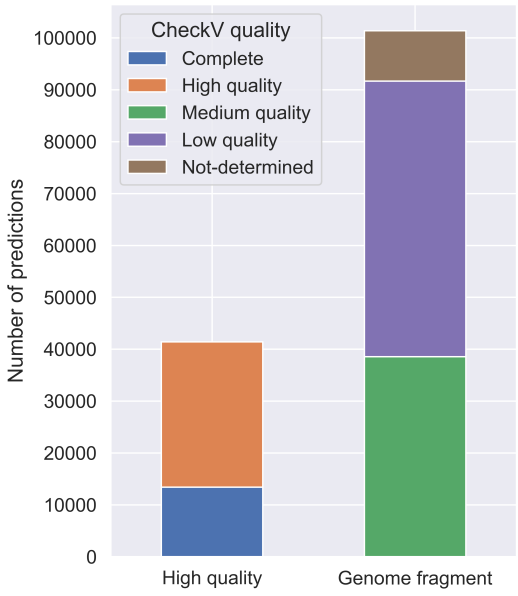
A


Genome size (kb)
100
90
80
70
60
50
40
30
20
10
0
GPD
IMG/VR
GVD
Database

B


Integrase
Holin
Reverse transcriptase
Terminal genes
Frequency
Terminase
Portal
Tail
Polymerase
Position across genome
Middle genes
Semi-terminal genes

C


CheckV quality
Complete
High quality
Medium quality
Low quality
Not-determined
Number of predictions
100000
90000
80000
70000
60000
50000
40000
30000
20000
10000
0
High quality
Genome fragment

D

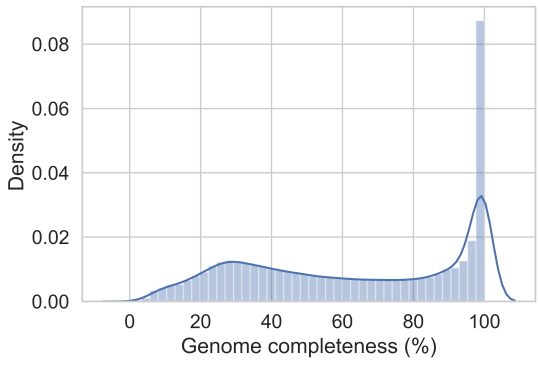

Density
0.08
0.06
0.04
0.02
0.00
0
20
40
60
80
100
Genome completeness (%)

Figure 3.4. Genome completeness of GPD. **A)** Compared to other public databases, GPD harbours the longest genomes with a median of 31 kb as opposed to 14 kb from IMG/VR and 11 kb from GVD. **B)** Distribution of phage marker genes across GPD predictions. Three main types of consensus distributions were observed, namely terminal, semi-terminal, and middle genes. **C)** Genome completeness as judged by CheckV. Over 40,000 genomes were categorized as high-quality (28%) (genome completeness > 90%), while the rest were predicted to be genome fragments. **D)** The median genome completeness of the whole database was was estimated to be 63.5%.

3.2.5 Clustering of phages into VCs

As explained above, I further organized the viral diversity contained in GPD into VCs. Even though a 95% nucleotide identity threshold has been proposed to delineate species in bacterial viruses (Adriaenssens and Brister, 2017), when I examined the final set of predictions (142,809), I realised that many phage genomes were still very similar between each other. Different predictions had extensive synteny with nucleotide identity < 95% and thus shared the majority of genes.

I then decided to explore further clustering by computing how many genomes were related to a "bait" genome at different thresholds of Mash distance (Figure 3.5A). Most of the genomes related to the bait were already saturating at a Mash distance of 10 (~90% nucleotide identity), which I considered as a more appropriate clustering threshold than a Mash distance of 5 (~95% nucleotide identity) (Figure 3.5B).

Since Mash doesn't take into consideration alignment fraction, I switched to BLAST to enforce a minimum alignment fraction of 75% of the shortest sequence and allowed a minimum of 90% nucleotide identity between genomes. In order to automatize the generation of clusters, I relied on an unsupervised approach, namely the Markov Clustering Algorithm or MCL (Dongen, 2000) (see Methods). In short, MCL uses random walks to automatically identify highly connected nodes (phage genomes in this case). After MCL clustering, GPD diversity ended up encapsulated in 21,012 non-singleton VCs. Benchmarking against the RefSeq phages revealed that GPD VCs were equivalent to a subgenus level, as >99% of all VCs were contained within a genus and in some cases, multiple VCs were associated to a single genus (Figure 3.5C).

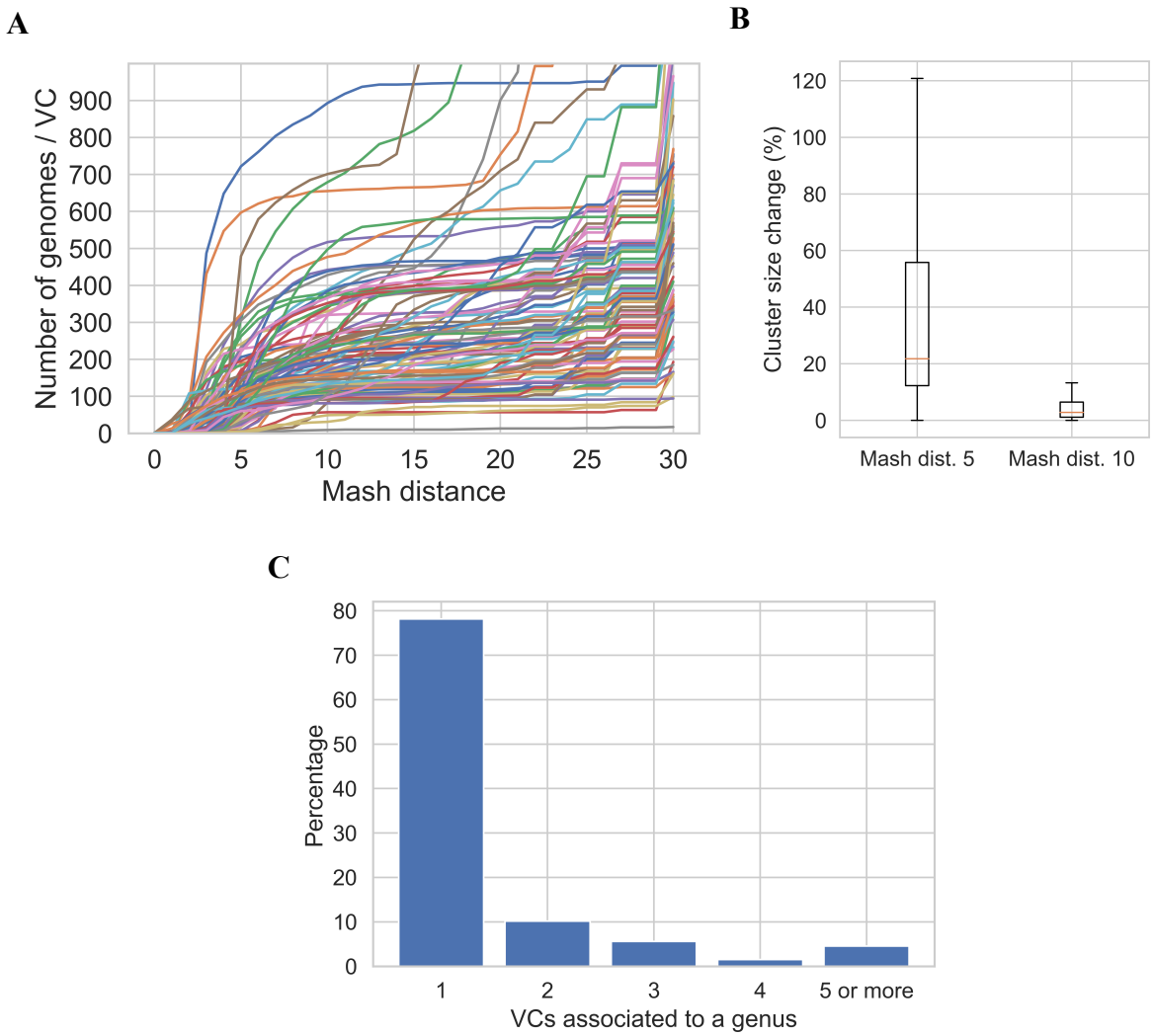


Figure 3.5. Clustering of phages into VCs. A) Even though 95% sequence similarity delineates species level in phages, I noticed extensive synteny between GPD predictions at that threshold. I explored other sequence identity thresholds by computing how many GPD genomes were related to a bait genome. **B)** Viral clusters started to saturate at a Mash distance of 10 (~90% sequence similarity), rather than 5 (~95% sequence similarity). **C)** Benchmarking against RefSeq phages showed that a single phage genus could be associated to several VCs, suggesting subgenus clustering.

3.2.6 Viral clusters reconstruct the phylogenetic structure of gut phages

The resultant VCs were not of uniform size but instead followed a negative exponential distribution with a few clusters (<50) composed of a large number of phage (>100 predictions) followed by a rapidly decreasing long tail of VCs with smaller membership size (Figure 3.6A).

This result suggested that genetic diversity is not evenly distributed in GPD. The number of genomes per VC could reflect inherent genetic diversity of a phage clade, however the most likely explanation here may be sampling bias (oversampled VCs will capture more genetic variation). The top VC was identified as the highly prevalent crAssphage (p-crAssphage), while the second contained a clade of phages characterized by a relatively long genome (~80kb), a BACON domain-containing protein, and *Bacteroidales* host range (hereafter referred to as the Gubaphage clade). The Gubaphage clade is a novel clade of gut phages proposed in this thesis and it is further characterized in Chapter 4. The phylogenetic structure of GPD could be visualized based on a network analysis of VCs (Figure 3.6B). Several VCs were highly inter-connected, forming super clusters and hinting to higher taxonomic clustering (e.g. viral subfamilies). On the other hand, isolated VCs may correspond to very genetically homogeneous viral clades.

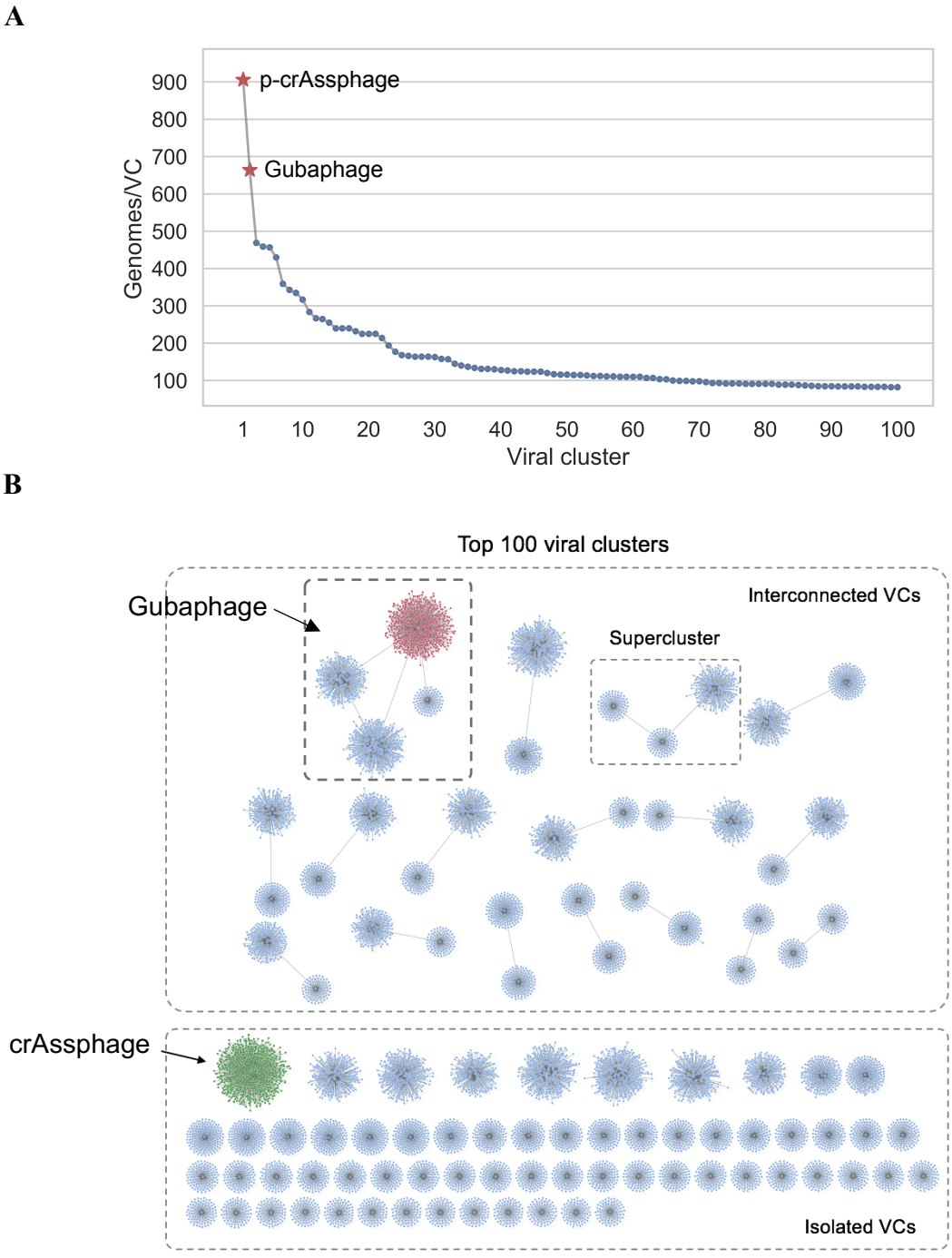


Figure 3.6. Distribution of genomes per VC and phylogenetic structure of GPD A) Distribution of genomes per VC. Only the 100 most prevalent VCs are shown. A member of the crAssphage family (p-crAssphage) was identified as the VC with the bigger cluster size, followed by a VC referred to as the Gubaphage. **B)** Visualization of the top 100 VCs reveal a subset of connected clusters and isolated ones. Inter-connection of VCs likely reflect higher phylogenetic structures such as subfamilies.

3.2.7 Bioinformatics tools

During the course of this work, I developed 3 bioinformatics tools that helped with the exploratory data analysis of GPD genomes, namely dotBlast (synteny analysis), hyperVir (visualization of hypervariable regions), and vMatch (classification of phage sequences). The development of these tools was motivated by the lack of ad-hoc bioinformatics tools to manage the sheer amount of genomes in GPD.

3.2.8 Synteny analysis for viral genomes (dotBlast)

During the exploratory analysis stage of this work I realised that I needed a high-throughput way to compare viral genomes. Sequence identity is a way forward, and adding coverage thresholds can lead to more robust strategies to assess similarity between two genomes. Nonetheless, the source of these two metrics (sequence identity and coverage) is the sequence alignment, and its inspection can help uncover more subtle differences such as insertions, deletions, and inversions.

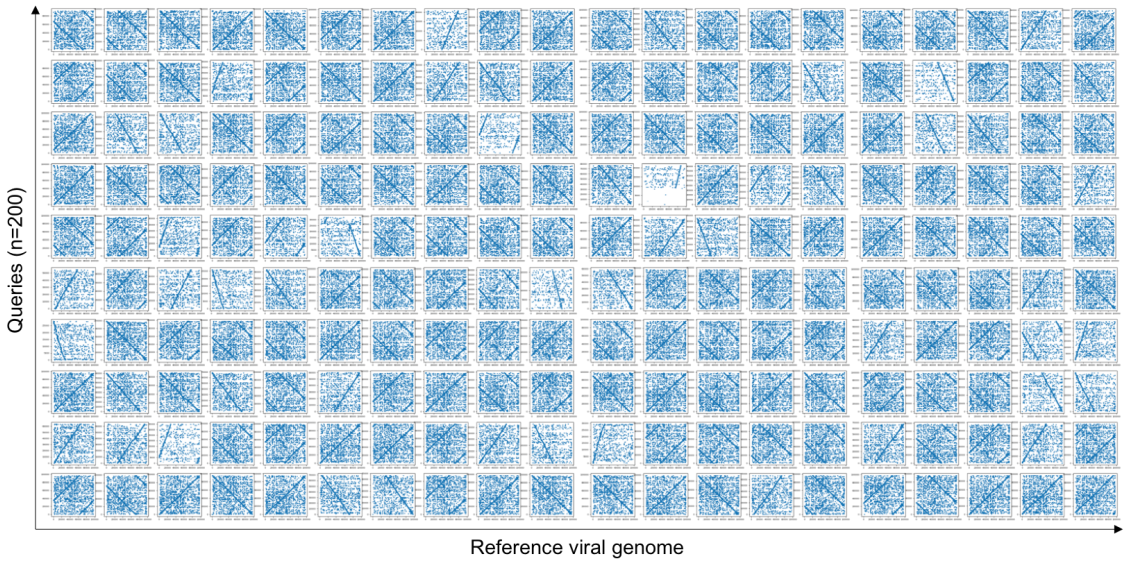
In bioinformatics, a dot plot (also known as a similarity matrix) is one way to efficiently visualize a pairwise sequence alignment. The dot plot was introduced in 1970 by Gibbs and McIntrye and it can be constructed by placing the bases of the first sequence as columns of a matrix, while the second sequence runs perpendicularly and thus fills up the rows of the matrix. Then we simply shade a cell in black if the residues in the corresponding column and row are identical. A consequence of this pattern is that matching subsequences appear as diagonal lines across the matrix.

If “n” and “m” are the lengths of the two sequences to analyse, then the number comparisons is n*m. However, generating the matrix this way is computationally inefficient (quadratic time complexity) and leads to a lot of noise. If a tool is meant to generate hundreds of dot plots in a reasonable amount of time, then this naïve strategy is not practical. A way around is simply to shade cells if they belong to a significant alignment. Fortunately, BLAST can readily process hundreds of queries in an efficient manner.

By incorporating the BLAST output of two aligned sequences, I developed dotBlast which given a blast reference viral genome and a set of queries, can quickly generate the coordinates for the generation of dot plots that compare each query to the reference (Figure 3.7A). In addition, in order to explore more conserved regions, the user can control the alignment significance threshold (Figure 3.7B). By generating dot plots, it's possible to have a quick glance of synteny across hundreds of queries against a reference (e.g. a member of a known viral subfamily). Analysis of dotplots can provide subtle details of genomic organisation e.g. a "broken" main diagonal may indicate circular genomes, a "jump" in the alignment can hint to an insertion or deletion.

With the increasingly large number of viral genomes mined from metagenomes, it is becoming more necessary to have high-throughput tools to easily visualize relationships between phage. DotBlast depends only on BLAST and Python, which are usually already available in a large number of bioinformatics systems or can be easily installed.

A



B

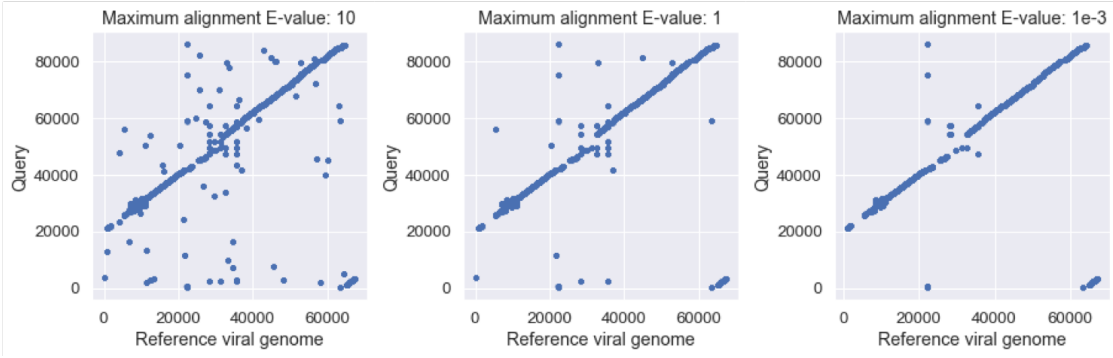


Figure 3.7. DotBlast tool. A) DotBlast can compare hundreds of viral genomes against a reference (e.g. a member of a viral subfamily) by generating dot plots. It uses BLAST to calculate significant alignments and plots them in a dot plot format in a fast manner. B) The significance of alignments can be controlled, allowing to identify highly conserved regions (or decrease noise).

3.2.9 Hypervariation analysis (hyperVir)

Having a large genetic diversity encapsulated in a clade of closely related viral genomes (e.g. species or genus) enables a large number of analyses. The discovery of hypervariation within proteins is particularly interesting because it can lead to the identification of genes with binding domains. These genes can be involved in recognition of bacterial receptors, binding of mucus, and even depolymerization of surface decorating polysaccharides by lytic phage enzymes. Analysis of gut viromes has suggested the existence of multiple hypervariable loci in gut phages (Minot et al., 2012), and thus the assessment of hypervariation in GPD phages can prove to be useful for their characterization. In order to facilitate hypervariation analysis in viral genomes I developed hyperVir which allows visualization of amino acid diversity and automatic detection of hypervariable regions in viral contigs.

The basic workflow (Figure 3.8A) involves an input FASTA file containing protein sequences, followed by a multiple sequence alignment with MAFFT, and finally the estimation of amino acid diversity at each position of the alignment by calculating Shannon's entropy. The signal is smoothed out by passing the Savitzky-Golay filter and hypervariable regions can be detected by a spike of amino acid diversity (Figure 3.8B).

HyperVir is thus a tool that conveniently can uncover viral genes with hypervariable domains which can help narrow down gene function. A more rigorous method involves the detection of positive selection with the Ka/Ks ratio. However, HyperVir is geared towards the detection of highly variable regions (hypervariation), speed, and high throughput visualization of results (Figure 3.8C).

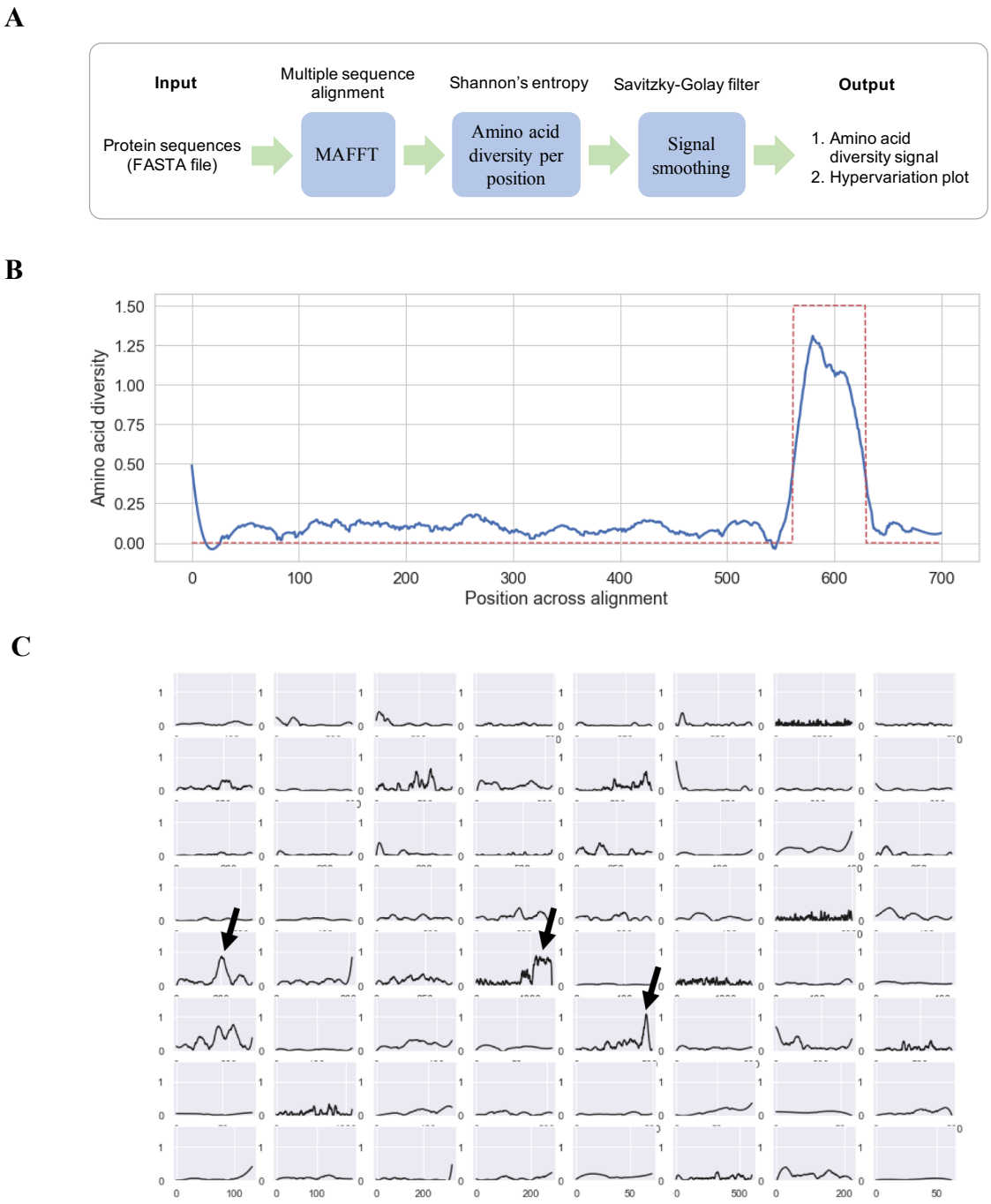


Figure 3.8. HyperVir tool. A) Pipeline to identify hypervariable genes. The input is a FASTA file containing a set proteins. After generating a multiple sequence alignment of the proteins, hyperVir calculates the amino acid diversity at each amino acid position by computing Shannon's entropy. Finally, the signal is smoothed with the Savitzky-Golay filter and the amino acid diversity plots visualized. **B)** Output of hyperVir. Amino acid variation is showed per position of the multiple sequence alignment. An hypervariable region is highlighted in red. **C)** hyperVir applied to 64 sets of proteins shows different hypervariation patterns. Pointed by arrows are examples of proteins with high hypervariation domains.

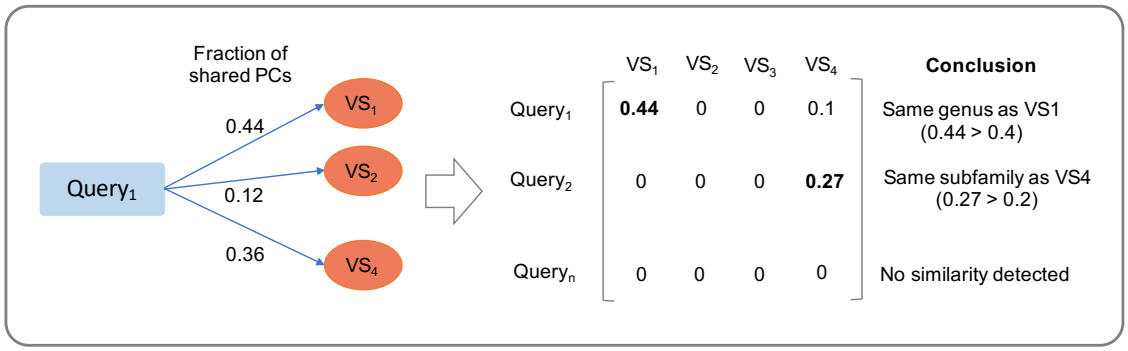
3.2.10 Exploring viral taxonomy through shared protein clusters (vMatch)

Large-scale classification of phage predictions is a recurrent challenge in metagenomic projects. Unlike bacteria, viruses lack a common marker gene and thus it's difficult to reliably estimate the phylogenetic distance between clades. This issue is compounded because phages often recombine and become mosaic, further blurring genetic distances between them. Finally, metagenomic projects often generate viral fragments which decrease the performance of methods that exploit specific-clade marker genes. The idea of using shared homologous proteins as a criterion to demarcate phage clades looked particularly promising e.g. the Phage Proteomic Tree (Rohwer and Edwards, 2002). In recent years, several tools were developed to harness the use of protein clusters to carry out phage taxonomy assignment. However, the majority of these methods were not implemented in packages, limiting their widespread use. A notable exception was the VICTOR tool, which was accessible online but had scalability issues (limit to 100 genomes) (Meier-Kolthoff and Göker, 2017). More recently, vContact2.0 combined a network approach with the idea of sharing protein clusters, and optimized it for the classification of viral predictions at the genus-level. Furthermore, vContact2.0 is also available as a standalone version, making it more accessible for custom datasets (Bin Jang et al., 2019).

Unfortunately, vContact2.0 is not scalable for huge datasets like GPD as the program could not finish processing the sheer volume of predictions (>140,000) submitted. Submission of shorter queries also failed to return taxonomy classification, but only the genus-like clusters. In addition, although useful, the genus scope of the program is a conservative taxonomy assignment. I believe that predictions can be more meaningfully placed into candidate viral subfamilies. This is particularly useful in metagenomes with huge novel viral diversity, as subfamilies can potentially bring together a multitude of novel genera that otherwise would be disconnected from known viral clades and deemed as "dark matter" of the dataset. Importantly, downstream analyses can be negatively affected, as hypothesis testing of associations of specific clades with another variable of interest (e.g. geographical distribution or disease) can end up underpowered. While the criteria for the inclusion of a phage into a specific viral subfamily varies, a sharing of at least 20% of homologous proteins between two genomes has been used to bioinformatically define viral subfamilies (Lavigne et al., 2008, 2009). This was the case of the crAss-like clade, in which the authors segregated all the crass-like sequences into viral subfamilies (20-40% sharing) and genera (>40% sharing) (Guerin et al., 2018).

With this in mind, my objective was to generate a tool for easy taxonomic exploratory data analysis of metagenomic datasets. I developed a standalone program (vMatch) for putative taxonomic assignment of metagenomic viral predictions against reference viral sequences (e.g. RefSeq) based on the principle of shared PCs to demarcate clades. vMatch takes in a file containing clusters of homologous proteins derived from pooling the proteome of the queries (e.g. metagenomic predictions) and reference viral sequences and then calculates the fraction of shared PCs between them. It then stores the results in a matrix in which the rows correspond to the queries and columns to the reference sequences (Figure 3.9A). Each entry corresponds to the pairwise mean of the shared PCs between the query and a reference. The matrix can then be visualized with a clustered heatmap. For instance, members of reference phage clades (*Skunavirus*, *Peduovirus*, *Pahexavirus*, *Teseptimavirus*) are columns of the heatmap, while rows are queries (Figure 3.9B). Clustering of the rows reveals a putative membership of the queries (e.g. metagenomic predictions). If the queries are also used as reference viral sequences, then visualization of the matrix enables the identification of novel clades (red boxes, Figure 3.9B).

A



B

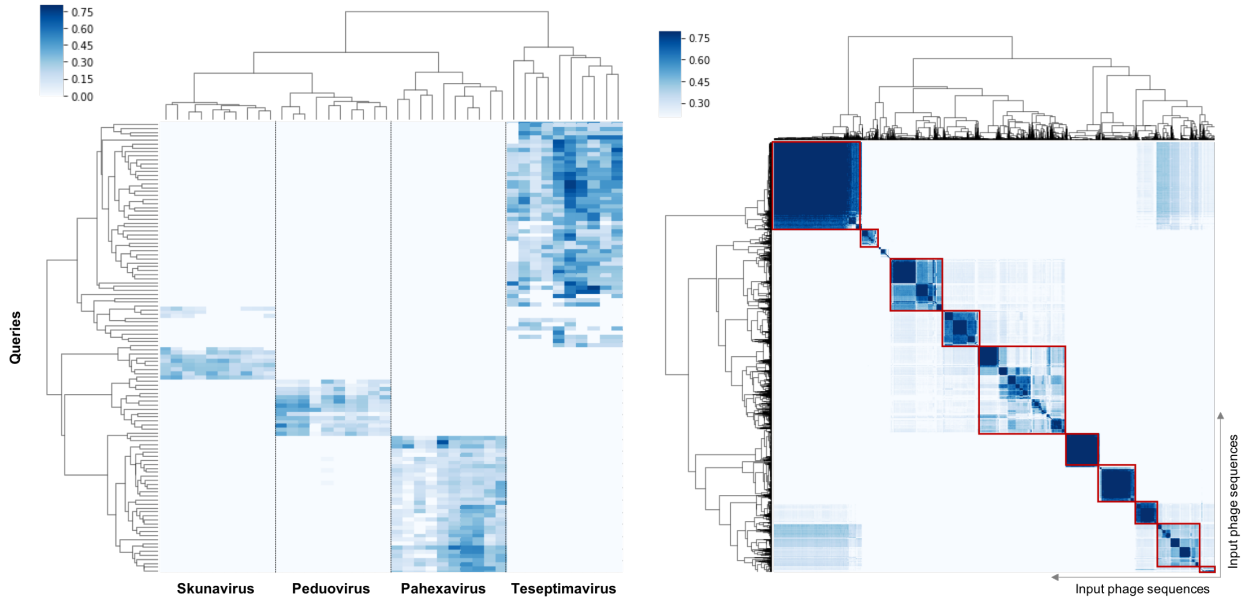


Figure 3.9. vMatch tool. A) Given a query and a set of viral sequences, vMatch calculates the fraction of shared protein clusters (PCs) between them as a proxy of their relationship. For instance, if two viral sequences share >20% of PCs, then they may belong to same candidate subfamily. **B)**. Visualization of vMatch results with clustered heatmap. On the left, a set of queries is compared against reference sequences, rows cluster according to their membership. On the right, the queries are also provided as reference sequences. The heatmap allows the easy identification of clades within the input sequences.

3.3 Conclusions

In this chapter, I presented the framework and rationale for the downstream analyses of human gut phages. By processing viral predictions from 28,060 gut metagenomes and 2898 bacterial isolate genomes, I generated a comprehensive and high-quality database of bacteriophage genomes, namely the gut phageome database (GPD). I showed that two popular tools for viral predictions (VirFinder and VirSorter) even with conservative settings, often predict integrative and conjugative elements (ICEs) as phages. I discovered that phages and ICEs significantly differ in gene density, fraction of hypothetical proteins, and kmer profile and thus these features can be exploited to segregate them. I trained a neural network to learn these differences and deployed it across thousands of predictions to minimize the number of false positives in GPD.

As reported in recent studies that analysed viromes from other environments, I uncovered an enormous amount of novel viral diversity in the human gut, which was particularly prominent when GPD is compared to the gold standard set of known viral genomes (RefSeq phages). This comparison highlighted three main things, namely the outstanding diversity of phages, the limited number of currently available high-quality phage genomes, and how mining of metagenomes can be harnessed to counter the lack of genomic data for phages. Comparing to other public phage databases, GPD outperformed in diversity and genome completeness by a wide margin. These improvements were due to the large number of metagenomes mined, and the diversity of samples which spanned all the 6 continents.

Even though viral predictions were non-redundant at 95% nucleotide identity (which roughly correspond to species level) (Adriaenssens and Brister, 2017), I noticed that at this threshold many predictions still had extensive synteny and nucleotide identity (>90%) to other predictions. For this reason, I decided to further group them into viral clusters (VCs) which consisted of more discrete viral populations. A recent study proposed to formalize the use of species-rank virus groups (Roux et al., 2019). This study found a cluster of genome pairs (suggestive of a species rank) that encompassed a large fraction of phage genomes with a nucleotide identity >90%, providing further support to a departure of the minimum 95% threshold. The generation of VCs is a powerful concept, because it enables to encapsulate highly related viruses into homogenous phage clades and allows to obtain better consensus of their inherent features such as their core and accessory genomes or average genome length. This becomes more evident in the next couple of chapters when I profile the biological

functions and epidemiology of gut phages. In addition, the quality of VCs defined in this work are benefited by the significantly longer genomes hosted by GPD (median>31kb), and provide more sensitivity to find distinctive features of a phage clade.

A critical step in this work was the exploratory data analysis. Unfortunately, none of the existing bioinformatic tools were suitable to handle the large number of GPD genomes. Thus, I decided to create standalone versions of programs that were useful during the development of this work. In addition, due to the large-scale nature of my dataset, processing speed was a priority and therefore all the tools are suitable for high-throughput analyses. The 3 programs developed here are suitable for the assessment of relatedness of viral genomes (dotBlast), study of hypervariation (hyperVir), and exploration of phage phylogeny by overlap of PCs (vMatch).

Chapter 4: Function, phylogeny and host assignment of gut phages

4.1 Introduction and aims

Analyses of predicted phage sequences from gut metagenomes have yielded fascinating insights into phage biology, such as the presence of sticky domains - which may facilitate adherence of some phage to the intestinal mucus (Barr et al., 2013) - reverse transcriptases to promote hypervariation (Minot et al., 2012), and proteins with ankyrin domains that may aid bacterial hosts in immune evasion (Jahn et al., 2019). However, previous functions have been inferred from bulk viral fragments, severely limiting the resolution to characterize individual phage genomes.

Due to the difficulty of culturing anaerobic gut bacteria, the identity of the hosts targeted by gut phages is a crucial but largely unanswered question. Often phages are restricted to infect single bacterial species, however distantly related gut bacteria have been found to harbour CRISPR spacers that target similar phages (Shkoporov et al., 2019) and almost identical prophages (Cornuault et al., 2018). These results suggest that gut phages may be more promiscuous than expected.

In this chapter, I describe common functions and auxiliary metabolic genes encoded by human gut bacteriophages. I also highlight instances of hypervariable domains which may indicate the presence of phage receptor binding proteins. I then shift the focus to the analysis of two clades of gut phages, namely the Gubaphage and the *Picovirinae* subfamily. The Gubaphage is the viral cluster (VC) with the highest number of GPD predictions after the p-crAssphage, while the *Picovirinae* subfamily was the most common predicted phage taxonomy in GPD. As I will show in Chapter 5, both clades are also highly prevalent across all continents. Finally, host assignment allows me to study patterns of phage diversity across bacterial clades of the human gut and investigate their host range patterns.

The aims of the research presented in this chapter are:

- uncover functions encoded by human gut bacteriophages;
- identify and characterize important phage clades of the human gut;
- carry out host assignment and investigate patterns of phage diversity across gut bacteria.

4.2 Results and discussion

4.2.1 Functions encoded by gut phages

Having a collection of over 142,000 viral genomes from the human gut allowed me to explore the functional patterns of gut bacteriophages at an unprecedented scale. In order to avoid biases due to a large number of highly genetically related viral predictions, I carried out the analysis at the level of VCs and ranked the results by fraction of VCs encoding the predicted functions. In addition, given that prophages are found in GPD predictions, I only considered regions classified as "viral" by checkV (Nayfach et al., 2020) to safeguard against bacterial DNA. I investigated the most ubiquitous KEGG pathways and modules encoded by gut phages (Figure 4.1A). The most frequent KEGG pathways detected were those associated with DNA replication (ko03030), mismatch repair (ko03430), purine and pyrimidine metabolism (ko00230, ko00240), homologous recombination (ko03440), and cysteine and methionine metabolism (ko00270). Although DNA replication, mismatch repair and homologous recombination can be thought of inherent pathways of phages, the last two are an example of auxiliary metabolic genes (AMGs). AMGs augment host metabolism during infection and have a bacterial origin (Breitbart et al., 2007). Inspection of purine and pyrimidine metabolism genes revealed that dUTPases and thymidylate synthases were prominent members of this category. Cellular dUTPases break down dUTP into dUMP and pyrophosphate, while thymidylate synthases convert dUMP into dTTP (Hizi and Herzig, 2015). Since most DNA polymerases can use dUTP instead of dTTP for DNA synthesis, gut phages can minimize the risk of misincorporation of uracil in their genome by lowering the intracellular dUTP/dTTP ratio with dUTPases and thymidylate synthases.

I also found other frequent functions related to the metabolism of sulphur-containing compounds such as assimilatory and dissimilatory sulphate reduction (M00176 and M00596). I decided to specifically search for hits that included the phosphoadenosine phosphosulfate reductase and sulfate adenylyltransferase as both enzymes participate in the reduction of sulfate (Muyzer and Stams, 2008). Sulfate reduction can be harnessed for assimilatory (anabolic) reactions which are involved in the biosynthesis of S-containing amino acids, as well as for dissimilatory pathways (energy generation) which use sulphur instead of molecular oxygen as an electron acceptor. This analysis unveiled 215 VCs that primarily infect *Bacteroides*,

Bacteroides B, *Parabacteroides*, *Prevotella*, *Bacteroides A*, and *Blautia_A*. Phages encoding sulphur metabolism enzymes may seem enigmatic, however dissimilatory reactions could be exploited by phages to ensure sustained energy generation in the gut anaerobic environment. For instance, cyanophages can encode photosynthetic genes in order to boost energy production during the infection stage (Clokie and Mann, 2006). Sulphur metabolism genes have also been found in dsDNA phages from the deep ocean, where it has been hypothesized that they may be involved in supplementing or sustaining sulphur oxidation metabolism in bacteria to ensure continued viral infection and replication (Anantharaman et al., 2014). While the top predicted hosts are not considered sulphur-reducing gut bacteria, it has been shown that *Parabacteroides* and *Bacteroides* isolated from chicken cecum express proteins related to sulfate assimilation. In addition, when dietary carbohydrates are scarce, *Bacteroides thetaiotaomicron* can degrade host glycans (heparin and heparin sulfate) which have variable sulfation patterns. *Prevotella* strain RS2 and *Bacteroides fragilis* are also considered mucin-degrading bacteria (Tailford et al., 2015). Thus, it remains a possibility that as these bacteria can metabolize sulphated compounds, phages could exploit sulphur pathways for their own advantage.

When I was inspecting annotations of individual genomes of GPD phages, I discovered multiple genes annotated as transporters. Therefore, I decided to quantify the most common phage transporters found in GPD (Figure 4.1B). Top hits corresponded to transporters for pantothenate, Zinc, Cobalt, Taurine, Nicotinamide mononucleotide, Nicotinamide riboside, spermidine/putrescine, and potassium.

Nutrient transporters have been identified in other phages. For instance, viral genomes from the North Atlantic Subtropical Gyre can code for the pstS gene which transports phosphate into the host (Warwick-Dugdale et al., 2019). Phosphate is a primary limiting nutrient in marine environments, so phages can benefit their host by coding for phosphate transporters. Certainly, phages isolated from phosphate limited environments have been found to carry more AMGs related to phosphate uptake than those from phosphate replete environments (Kelly et al., 2013). It's known that the human gut is not a homogenous environment but one with nutrients that vary in space and time (gut biogeography) (Donaldson et al., 2016). Thus, the type of transporters coded by phages may depend on nutrients that maximize the chances of survival of their bacterial host at a specific gut niche. In line with this thought, substrates that aid anaerobic respiration may be more common in the most hypoxic areas of the gut such as the

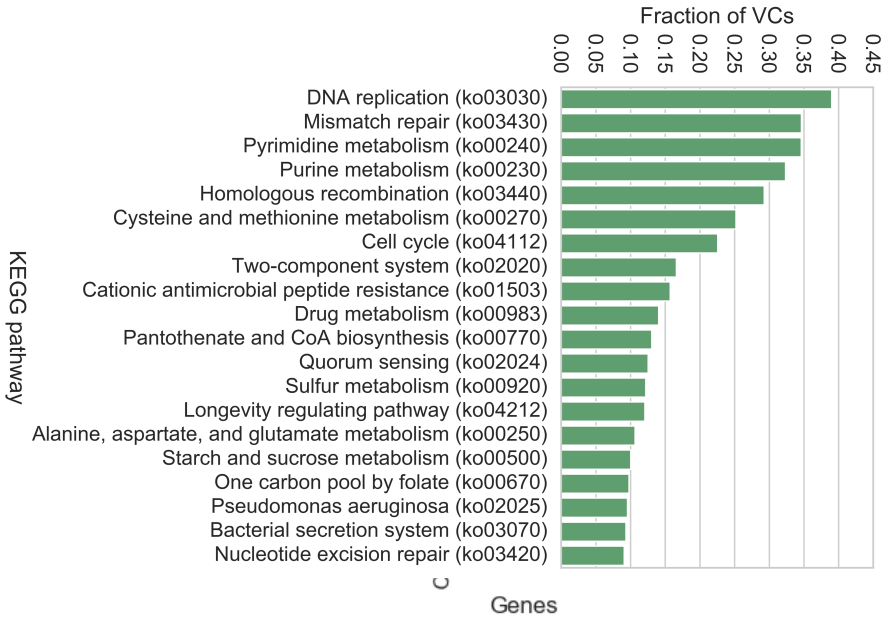
large intestine. For instance, Taurine (a major constituent of bile) can be metabolized into sulfite, enabling anaerobic respiration. Small amounts of bile salts that were not absorbed in the small intestine, may be better harnessed by phages coding for taurine transporters in the hypoxic environment of the large intestine.

I then shifted my attention to investigate the incidence of specific genes previously found in viral metagenomes from human faeces such as reverse transcriptases (Minot et al., 2012) and sticky domains (Barr et al., 2013).

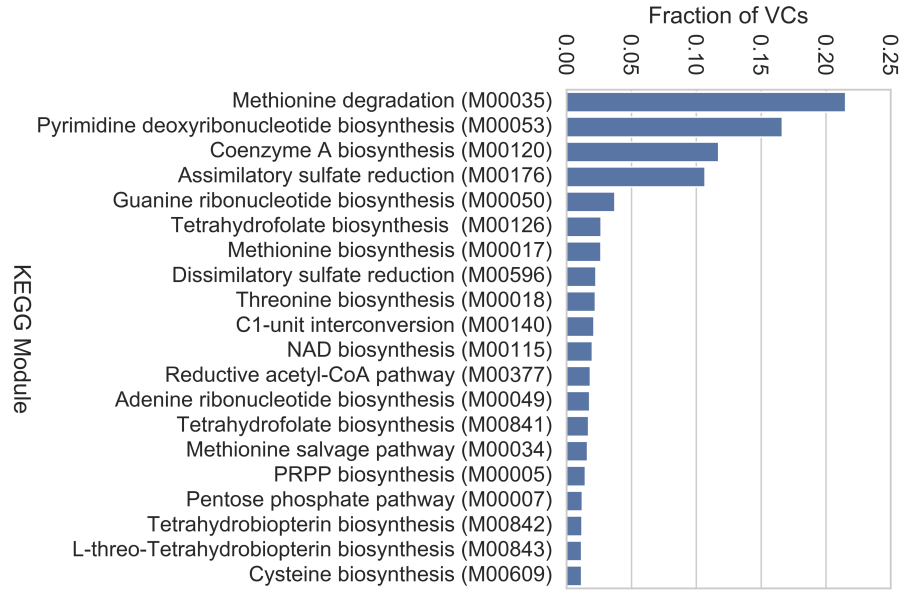
Over 2500 VCs (~12% of all VCs) encode reverse transcriptases (RTs) (Figure 4.1C). RTs in phages have been found to play a role in the generation of sequence diversity in target phage genes such as receptor binding proteins, and thus RTs with that function are called diversity-generating retroelements (Liu et al., 2002). The high incidence of RTs found here contrasts with previous reports that found very low prevalence of DGRs in phages (3 phages in ~600 dsDNA phages from NCBI) (Schillinger and Zingler, 2012). Similarly, When I analysed the incidence of RTs in RefSeq phages, only 0.38% contained them. Recently, it was reported that retrons, which are composed of a RT and a non-coding RNA, can work as an anti-phage defence system (Millman et al., 2020) . It's possible that many RTs carried by gut phages may be involved in defending against other phages, thus providing their host a selective advantage.

I also detected phage genes with adhesive domains (Figure 4.1C). For instance, Immunoglobulin-like (Ig-like) domains which occur frequently on the surface of the *Caudovirales* (Fraser et al., 2006), were found in ~5% of VCs. The Bacteroides-Associated Carbohydrate-Binding Often N-terminal domain (BACON), which has been hypothesized to help phages bind intestinal mucin (de Jonge et al., 2019), was found in 0.88% of VCs. Finally, the collagen triple helix repeat (CTHR) was found in ~8% of VCs. Collagens domains have been suggested to aid in the attachment of phages to *E.* coli (Yu et al., 2014). Sticky domains in phages are often found close to tail genes, and it has been suggested that they may facilitate phage adsorption to its host (Fokine and Rossmann, 2014). In many cases, successful phage infections in the gut are mediated by the correct combination of sticky domains and capsular polysaccharides on the surface of bacteria (Porter et al., 2020).

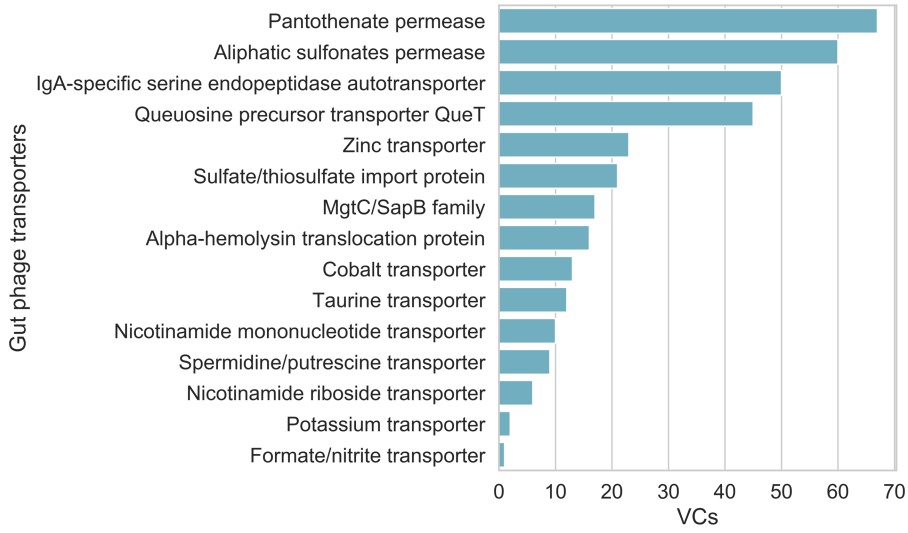
A


Fraction of VCs
0.00
0.05
0.10
0.15
0.20
0.25
0.30
0.35
0.40
0.45
KEGG pathway
DNA replication (ko03030)
Mismatch repair (ko03430)
Pyrimidine metabolism (ko00240)
Purine metabolism (ko00230)
Homologous recombination (ko03440)
Cysteine and methionine metabolism (ko00270)
Cell cycle (ko04112)
Two-component system (ko02020)
Cationic antimicrobial peptide resistance (ko01503)
Drug metabolism (ko00983)
Pantothenate and CoA biosynthesis (ko00770)
Quorum sensing (ko02024)
Sulfur metabolism (ko00920)
Longevity regulating pathway (ko04212)
Alanine, aspartate, and glutamate metabolism (ko00250)
Starch and sucrose metabolism (ko00500)
One carbon pool by folate (ko00670)
Pseudomonas aeruginosa (ko02025)
Bacterial secretion system (ko03070)
Nucleotide excision repair (ko03420)
Genes

B


Fraction of VCs
0.00
0.05
0.10
0.15
0.20
0.25
KEGG Module
Methionine degradation (M00035)
Pyrimidine deoxyribonucleotide biosynthesis (M00053)
Coenzyme A biosynthesis (M00120)
Assimilatory sulfate reduction (M00176)
Guanine ribonucleotide biosynthesis (M00050)
Tetrahydrofolate biosynthesis (M00126)
Methionine biosynthesis (M00017)
Dissimilatory sulfate reduction (M00596)
Threonine biosynthesis (M00018)
C1-unit interconversion (M00140)
NAD biosynthesis (M00115)
Reductive acetyl-CoA pathway (M00377)
Adenine ribonucleotide biosynthesis (M00049)
Tetrahydrofolate biosynthesis (M00841)
Methionine salvage pathway (M00034)
PRPP biosynthesis (M00005)
Pentose phosphate pathway (M00007)
Tetrahydrobiopterin biosynthesis (M00842)
L-threo-Tetrahydrobiopterin biosynthesis (M00843)
Cysteine biosynthesis (M00609)

C


Gut phage transporters
Pantothenate permease
Aliphatic sulfonates permease
IgA-specific serine endopeptidase autotransporter
Queuosine precursor transporter QueT
Zinc transporter
Sulfate/thiosulfate import protein
MgtC/SapB family
Alpha-hemolysin translocation protein
Cobalt transporter
Taurine transporter
Nicotinamide mononucleotide transporter
Spermidine/putrescine transporter
Nicotinamide riboside transporter
Potassium transporter
Formate/nitrite transporter
0
10
20
30
40
50
60
70
VCs

D

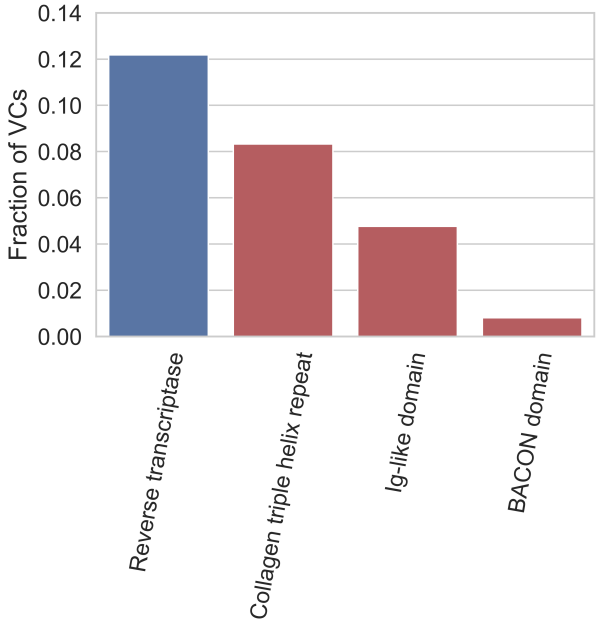


Figure 4.1. Functions encoded by gut phages. **A)** Top functions encoded by gut phages. Common functions included KEGG pathways and modules related to DNA replication and DNA repair. However, I also detected instances of auxiliary metabolic functions such as those involved in nucleotide and sulphur metabolism. **B)** Transporters found in gut phages which may provide a selective advantage to their hosts depending on its intestinal niche. **C)** Reverse transcriptases (RTs) can help phages to generate sequence diversity and potentially act as defence systems against other phages. Sticky domains (red) may facilitate adsorption to hosts and binding to intestinal mucus.

4.2.2 Protein clusters encoded by gut phages

While the functions described above corresponded to curated pathways and targeted searches, I then took a more agnostic approach by analysing the whole proteome of GPD. I clustered all the GPD proteins with the phage RefSeq proteome to understand the functions encoded by the resultant protein clusters (PCs) (Figure 4.2A). After removing singletons I ended up with 172,449 PCs. Top hits included PCs containing proteins involved in the integration of DNA into the host and the maintenance of a lysogenic state (anti-repressor and integrases), DNA processing (single-stranded DNA-binding protein), pore formation for DNA injection (tape-measure protein), DNA packaging into procapsids (terminases), and DNA methylases (defence against host endonucleases). Interestingly, the 11th most common PC (PC_11) which was

encoded by ~8.5% of all VCs could not be clustered with any viral protein from RefSeq. I inferred that this PC encompassed a family of relatively large (median: 259 aa, IQR: 33 aa) single-pass membrane proteins, as they carry a transmembrane region near the N-terminus. Submission of members of PC11 to HHpred (Söding et al., 2005), one of the most sensitive tools for protein homology detection, could not retrieve confident hits. Prediction of the host range of phages carrying proteins that belonged to this PC11, showed that it was mainly found in the Firmicutes phyla. This unknown PC highlights our lack of understanding of 'core' phage proteins that are widely spread in phages.

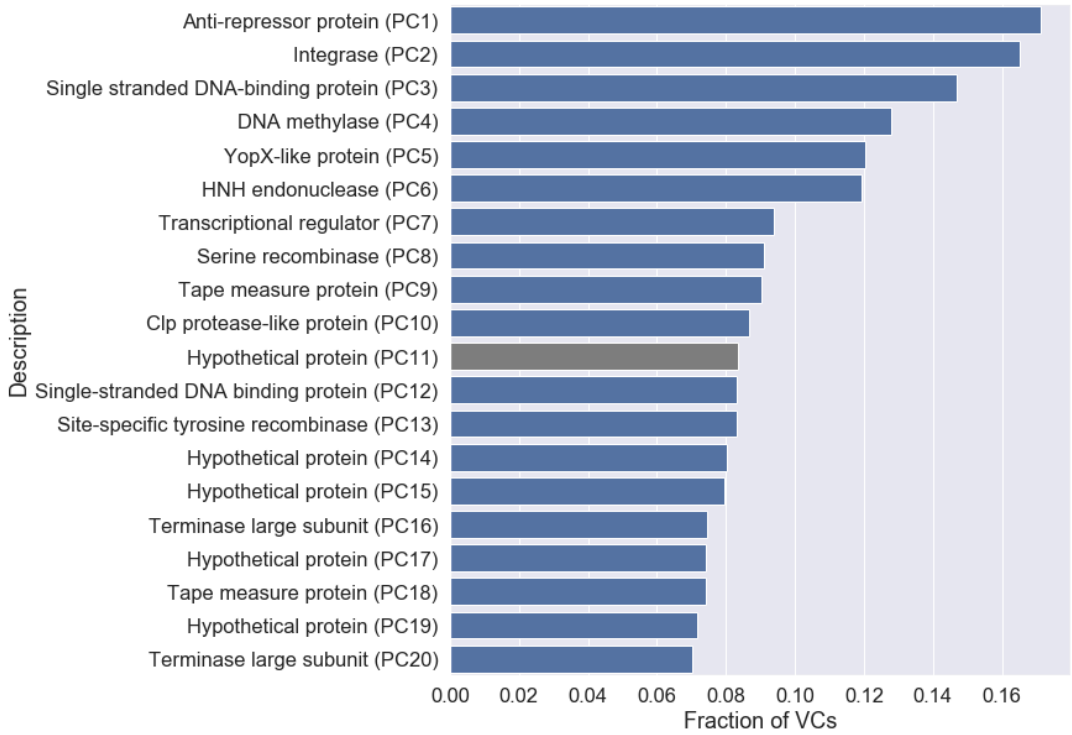


Figure 4.2. Protein clusters (PCs) encoded by gut phages. Prediction of the whole proteome found in GPD and RefSeq phages resulted in the generation of 172,449 PCs. After ranking the PCs by fraction of VCs they were encoded in, the top hits corresponded viral functions such as anti-repressor proteins, integrases, and structural proteins. Interestingly, one of the PCs found in ~8% of the VCs could not be assigned a function based on RefSeq proteins.

4.2.3 Identification of hypervariation domains uncovers putative phage tropism determinants

Prediction of the gene that confers bacterial host specificity to a phage (receptor binding protein) is important for characterization purposes but also because it can be mutagenized to expand the host range (Dunne et al., 2019). The latter is particularly interesting as viruses with broad host range can be harnessed to improve the effectiveness of phage therapy against antibiotic resistant bacteria (Yehl et al., 2019). Receptor binding proteins (RBPs) recognize a bacterial membrane protein (phage receptor) which facilitates adsorption of the phage onto their host (Dowah and Clokie, 2018). As a countermeasure to avoid infection, bacteria often mutate their receptor. However, phages respond by evolving their RBPs to recognize the new receptor. This predator-prey dynamics give rise to hypervariation in the binding domain of the RBPs and the bacterial receptor (Hampton et al., 2020).

I exploited the genetic variation present in the top VC of GPD to identify a candidate RBP for p-crAssphage (Figure 4.3A). After clustering the whole proteome of the crAssphage VC at >70% sequence identity and >90% coverage of both sequences, I sought to quantify amino acid diversity along a cluster of homologous crAssphage proteins. A sudden surge in diversity (hypervariation) would indicate the presence of a binding domain involved in host recognition. I identified such pattern in a group of homologous proteins predicted to be tail fibres. Attachment of tailed phages to bacteria is often mediated by tail fibres and surface receptors, providing further evidence that this set of proteins represent the RBP of p-crAssphage. The spike of amino acid diversity spanned ~70 amino acids and was located at the C-terminus. This finding is consistent with other phage receptor binding proteins that have their hypervariable domains at the C-terminus (Dunne et al., 2019).

I repeated the same exercise but with genomes found in the VC which corresponds to the Gubaphage clade (Figure 4.3B). I identified a large protein (> 2000 amino acids) with a hypervariable region of ~150 amino acids. Proximal genes to this protein included the major capsid protein and the terminase which due to phage modularity tend to be close to tail genes, so the identified protein with an hypervariable domain from Gubaphage is well suited to be a candidate receptor binding protein.

Thus, identification of hypervariable regions can help narrow down the function of important phage genes such as their receptor binding proteins. Elucidation of alternative strategies to homology search can prove invaluable in the characterization of the large fraction of hypothetical proteins in phages.

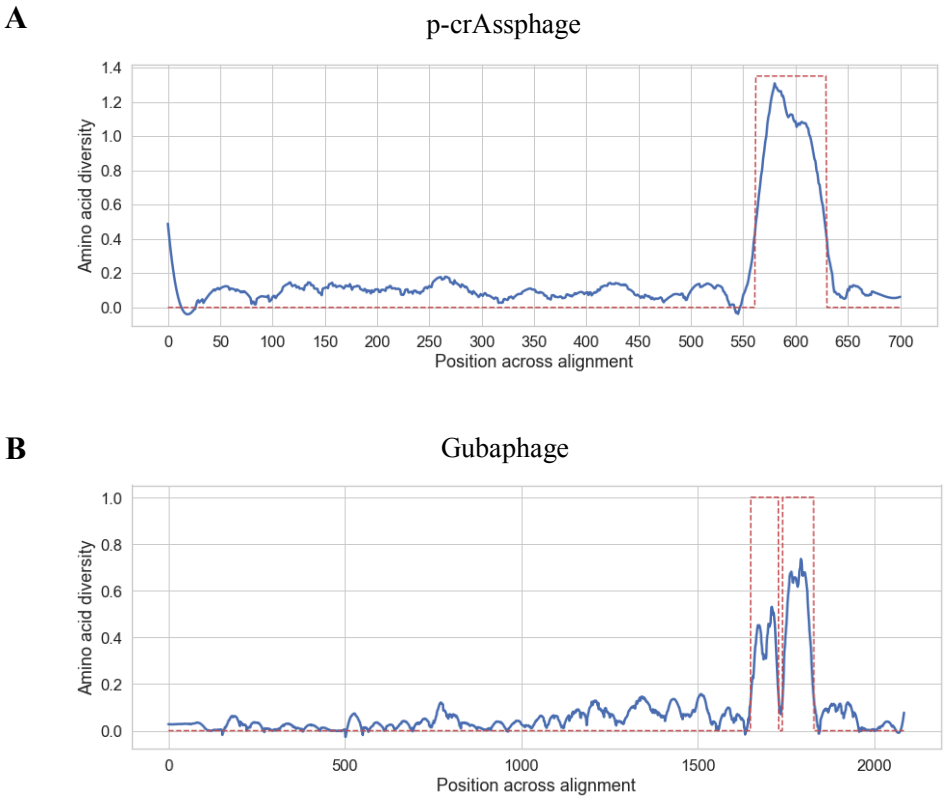


Figure 4.3. Hypervariable domains can narrow down protein function in phages. Detection of hypervariation protein domains can be useful to narrow down protein function in phages. Using this strategy I was able to identify candidate proteins to be the receptor binding proteins of the p-crAssphage **A)** and the Gubaphage clade **B)**.

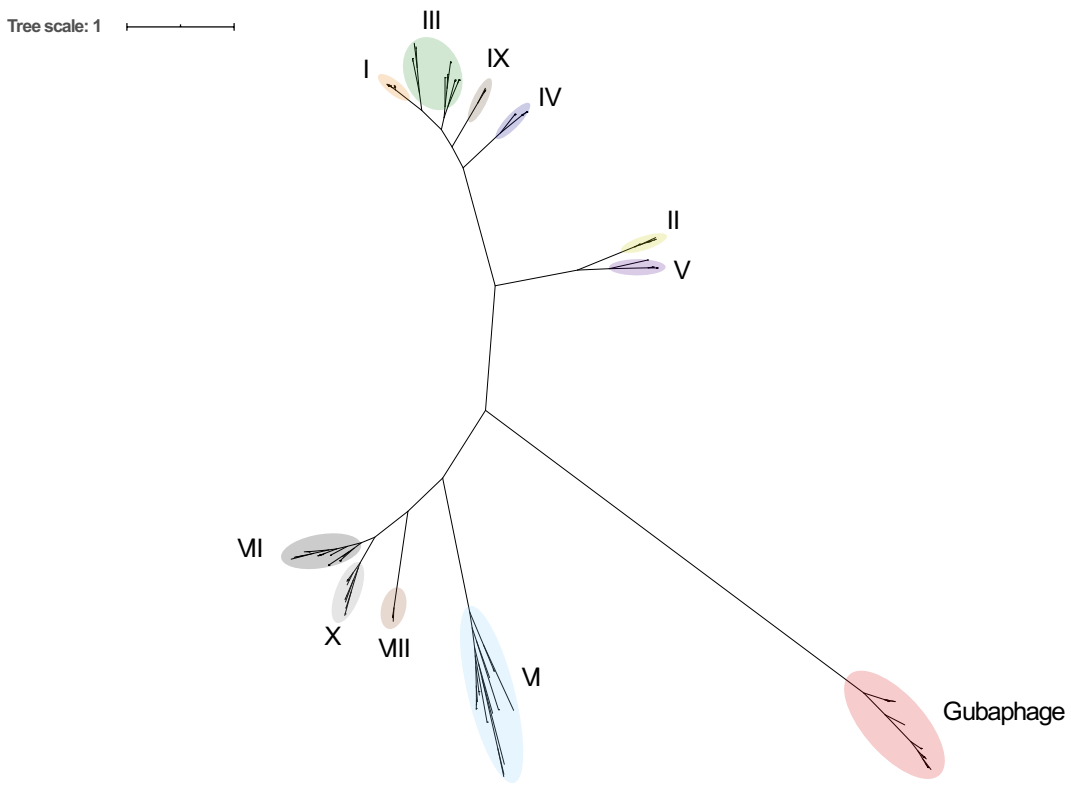
4.2.4 The Gubaphage represents a novel clade of gut phages

As mentioned in the previous chapter, the top two VCs of GPD predictions (p-crAssphage and Gubaphage) represented outliers regarding genetic diversity (as number of genomes / VC). Nucleotide sequence alignment with p-crAssphage revealed no significant similarity. However, they shared some functional features such as large genome size (>80 kb), a BACON domain-containing protein, predicted *Bacteroides* host range, and circular genomes. Searching

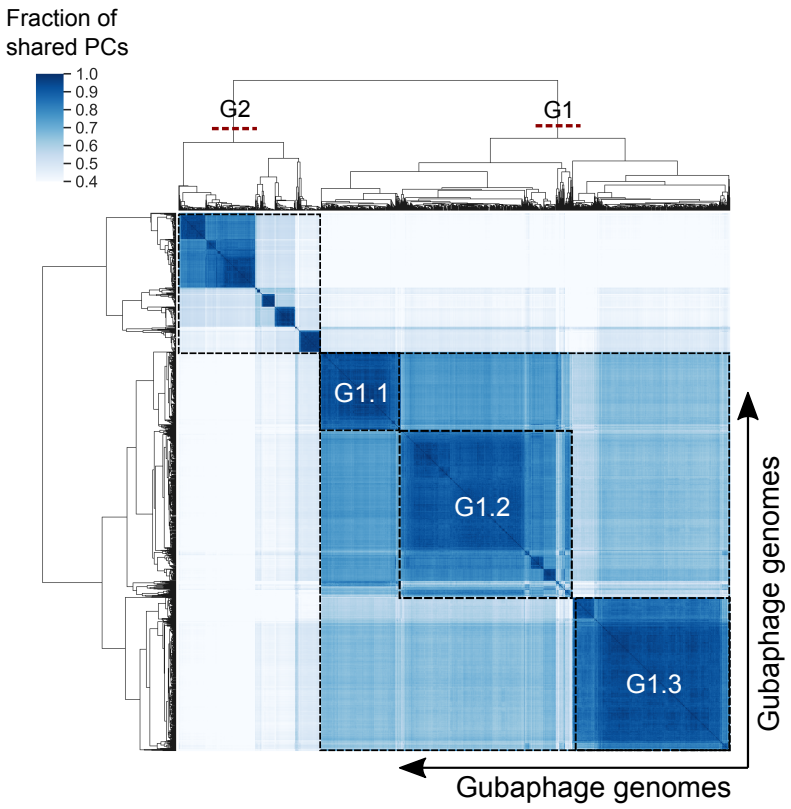
for sequences in the GPD with significant similarity to the Gubaphage large terminase gene (E-value < $1x10^{-6}$), I identified other 205 related VCs. Given its reminiscent features to crAssphage, I decided to investigate if the Gubaphage belonged to the recently proposed crAss-like family which consists of 10 genera and 4 subfamilies (Guerin et al., 2018). I examined this relationship by building a phylogenetic tree using the large terminase gene (Figure 4.4A). The tree successfully clustered all the crAss-like genera as expected, however the Gubaphage significantly diverged from the other crAss-like phages forming a distinct clade.

I then sought to characterize the phylogenetic structure of Gubaphage (Figure 4.4B). Analysis of protein overlap between Gubaphage's genomes revealed that this clade is composed of 2 clusters that share more than 20% but less than 40% of homologous proteins between them. This structure suggests two genera (G1 and G2) from a single viral subfamily. In addition, within G1 I identified another phylogenetic substructure composed of 3 large clusters (G1.1, G1.2, and G1.3) composed of 313, 514, and 502 phage genomes respectively. Host range prediction revealed that G1.1 infects *Bacteroides caccae* and *Bacteroides xylanisolvens* B, G1.3 *Bacteroides* B *vulgatus*, and G2 *Parabacteroides merdae* and *Parabacteroides distasonis*. In the case of G1.2, I couldn't confidently predict a putative host. Interestingly, the larger genetic distance between G1 and G2 also resulted in a more extreme host range switch, from Bacteroidaceae (G1) to Porphyromonadaceae (G2). Core genes of the Gubaphage included homing endonucleases, DNA polymerase I, FluMu terminase, DNA primase, DNA helicase, Thymidylate kinase, dUTPase, among others. Annotation of its genome revealed that Gubaphage is organized into three distinct regions (Figure 4.4C). One region encodes DNA machinery, the second is composed mainly structural genes and the third codes for a series of hypothetical proteins.

A


Tree scale: 1
III
I
IX
IV
II
V
VII
X
VIII
VI
Gubaphage

B


Fraction of shared PCs
1.0
0.9
0.8
0.7
0.6
0.5
0.4
G2
G1
G1.1
G1.2
G1.3
Gubaphage genomes
Gubaphage genomes

C

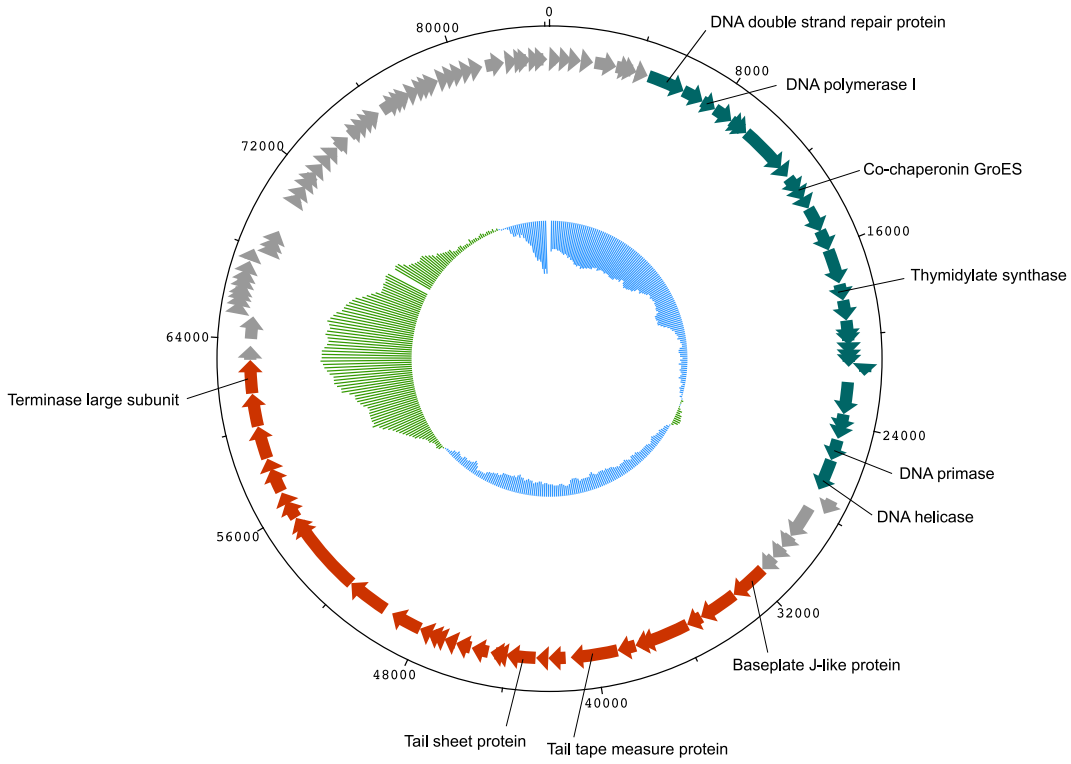


Figure 4.4. The Gubaphage clade. A) Unrooted tree showing the relationship of the crass-like phages and the Gubaphage. Each of the crAss-like clades (I to X), represents a different genus. The Gubaphage forms a clade of its own, suggesting a distant relationship to the crAss-like phages. The tree was constructed by carrying out a multiple alignment of the large terminase genes. **B)** Analysis of Gubaphage phylogenetic structure revealed two genera infecting member of the *Bacteroides* (G1) and *Parabacteroides* (G2) genera. **C)** Inspection of Gubaphage genome reveals that it is composed of 3 parts. The first one (blue-green) codes for DNA machinery, the second (red) harbours structural proteins such as the large terminase, and tail proteins, the third (grey top left) consists of only hypothetical proteins. Inner bars represents GC skew.

4.2.5 Expansion of the *Picovirinae* subfamily

Hitherto I have focused on novel phage clades (crAss-like family and Gubaphage clade), however phages belonging to traditional phage subfamilies such as *Spounavirinae*, *Peduovirinae*, *Autographivirinae*, and *Picovirinae* have been detected in human faces (Waller

et al., 2014). I decided to explore the diversity of the *Picovirinae* subfamily because it was one of the most common taxa predicted in GPD.

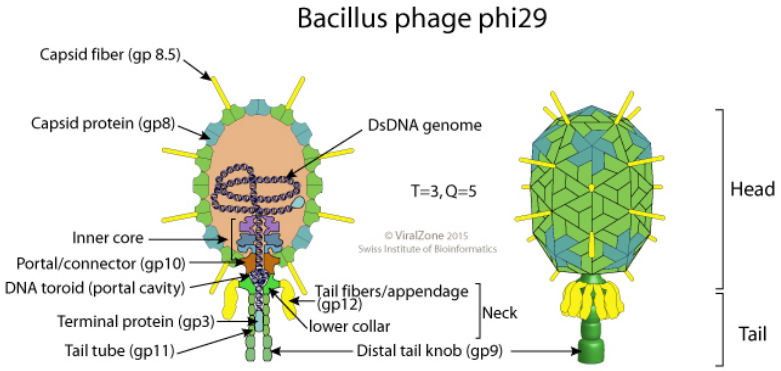
Picovirinae phages are known to have a small linear double stranded DNA genome of about 16-20 kb. They belong to the *Caudovirales* order and have an icosahedral capsid with a non-contractile tail (Figure 4.5A). The *Picovirinae* subfamily is currently composed of 3 genera namely *Salasvirus*, *Negarvirus*, and *Cepanuvirus* (Hulo et al., 2011). I predicted all the phages in GPD from this family by using a marker gene approach and obtained 4807 genomes.

In order to study the phylogenetic structure of the recovered genomes, I calculated all the pairwise overlaps of protein clusters between the *Picovirinae* genomes. Interestingly, after clustering the genomes and visualizing them in a heatmap, a phylogenetic substructure consisting of 4 large clades emerged (Figure 4.5B). Furthermore, an unrooted tree inferred from the PCs overlap clearly suggested 4 clades (Figure 4.5C). Given this evidence, I decided to structure the *Picovirinae* subfamily into 4 clades: Picovirinae_1 (P1), Picovirinae_2 (P2), Picovirinae_3 (P3), and Picovirinae_4 (P4). In addition, P1 clade was clearly divided into two clades, Picovirinae_1_1 (P1_1) and Picovirinae_1_2 (P1_2). With this new structure I was able to assign a clade to the three classified genera, while *Salasvirus* were assigned to P2, *Cepanuvirus* and *Negarvirus* were assigned to P1_1. In addition, I assigned a clade to several unclassified members of the *Picovirinae* with this expanded phylogenetic structure. Notably, P1_2, P3, and P4 remained without any known *Picovirinae* phage members assigned to them.

Host assignment revealed more than 288 gut bacteria isolates distributed between the Firmicutes and Actinobacteriota, moreover, P1_2, P3 and P4 were restricted to the Firmicutes, leaving P1_1 as the only inter-phyla *Picovirinae* clade. Containment of phage clades to a specific phylum is expected, as very distantly related host bacteria can present challenges to polyvalent phages e.g. substantially different replication machinery. In total, 31 genera of the human gut microbiota were predicted to be susceptible to infection by *Picovirinae* phages (Figure 4.5D).

This finding represents a clear example of the importance of metagenomics to fill in viral diversity gaps. In addition, gaining further knowledge of *Picovirinae* phages is important because their lytic lifestyle is suitable for phage therapy directed to Actinobacteriota and the Firmicutes.

A



B

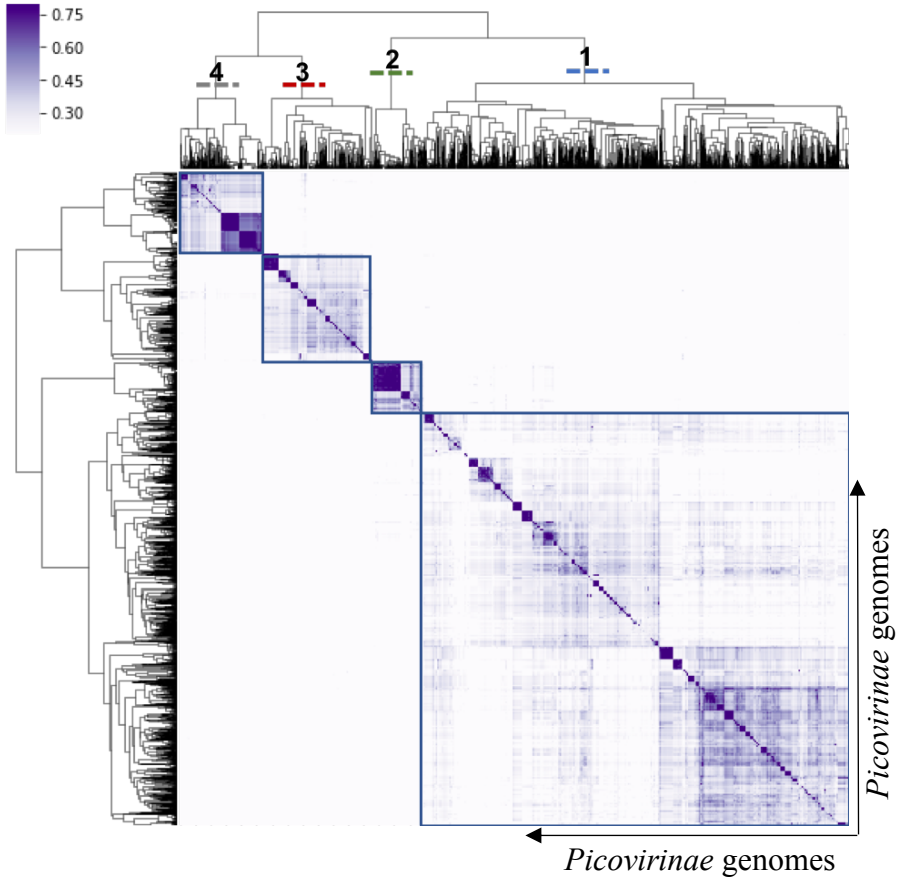
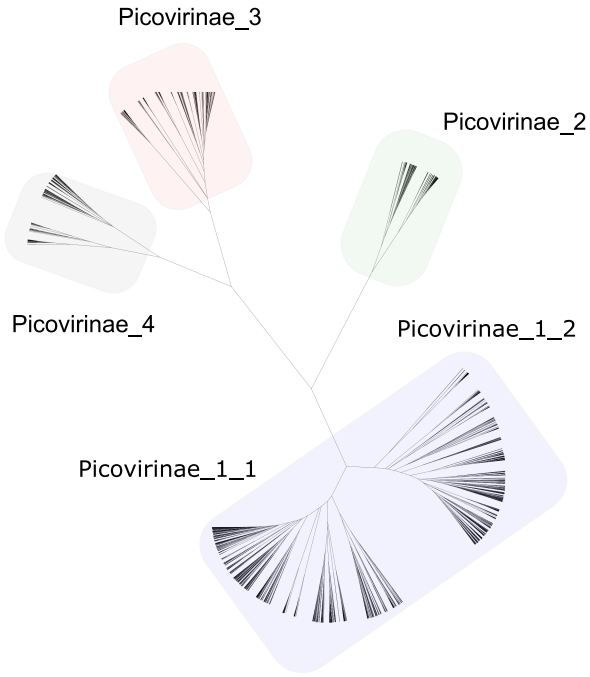


Figure 4.5. Expansion of the *Picovirinae* subfamily. A) The *Picovirinae* subfamily is characterized by having relatively small genomes (16-20kb) and a lytic lifecycle. They possess a linear double stranded DNA and have an icosahedral capsid with a non-contractile tail. **B)** Analysis of the phylogenetic structure of gut *Picovirinae* phages by fraction of shared protein clusters suggested 4 large clades.

C



D

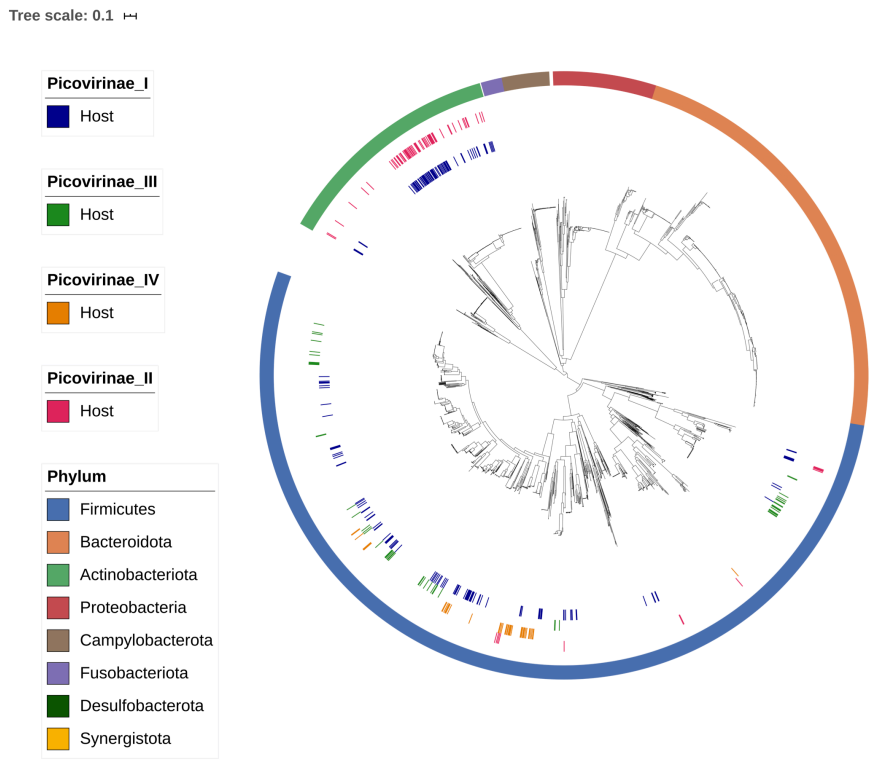


Figure 4.5. Expansion of the *Picovirinae* subfamily. C) Unrooted tree of shared protein clusters. The 4 clades were named Picovirinae_1, Picovirinae_2, Picovirinae_3, Picovirinae_4. This expanded diversity of the *Picovirinae* was able to accommodate the 3 known genera and several unclassified phages. Notably, Picovirinae 3 and 4 represented completely novel clades.

The tree was generated by calculating the fraction of shared protein clusters among individual Picovirinae phages and then carrying out hierarchical clustering with average linkage and Euclidean metric. **D)** Host assignment of *Picovirinae* phages to gut bacteria. Hosts were predicted by CRISPR spacer exact matching and prophage assignment. The tree was built by concatenating 40 universal core marker genes from each of the 2898 gut bacteria isolates and then carrying out a multiple sequence alignment. P1_2, P3 and P4 were restricted to the Firmicutes, leaving P1_1 as the only inter-phyla *Picovirinae* clade (Firmicutes and Actinobacteriota host range).

4.2.6 Viral diversity across gut bacteria clades

I next inferred the most likely bacterial hosts for each phage prediction using a comprehensive collection of 2898 human gut microbiota isolate genomes. By screening for the presence of CRISPR spacers (Edwards et al., 2016) targeting phage and by linking the prophages to their assemblies of origin, I was able to carry out host assignment. In order to estimate the rate of false positives (FPs) due to CRISPR random matches, I generated synthetic random spacers and mapped them against the GPD. Repeating this procedure 100 times revealed the distribution of the expected number of FPs across different matching criteria (Figure 4.6A). As can be seen from the graphs, no FPs are detected due to random chance when no mismatches are allowed across the whole length of the spacer (the criteria used in this work for the original mapping). However, as more mismatches are allowed, there is an increase in random matches across all coverages tested. Notably, at 80% coverage and only 4 mismatches allowed, the expected false positive rate due to random chance reach 2.6% of all the matches reported from the original mapping.

In total, I assigned 2,157 hosts to 40,932 GPD phage (28.66% of all predictions). This corresponded to at least one phage for 74.43% of all cultured human gut bacteria. I then analysed if there was any preference for phage infection across 5 common human gut bacterial phyla (Firmicutes, Bacteroides, Proteobacteria, and Actinobacteriota). At the phylum level, I detected significant lower phage prevalence in Actinobacteriota, with 58.79% infected isolates compared to at least 70% for the other phyla (Figure 4.6B).

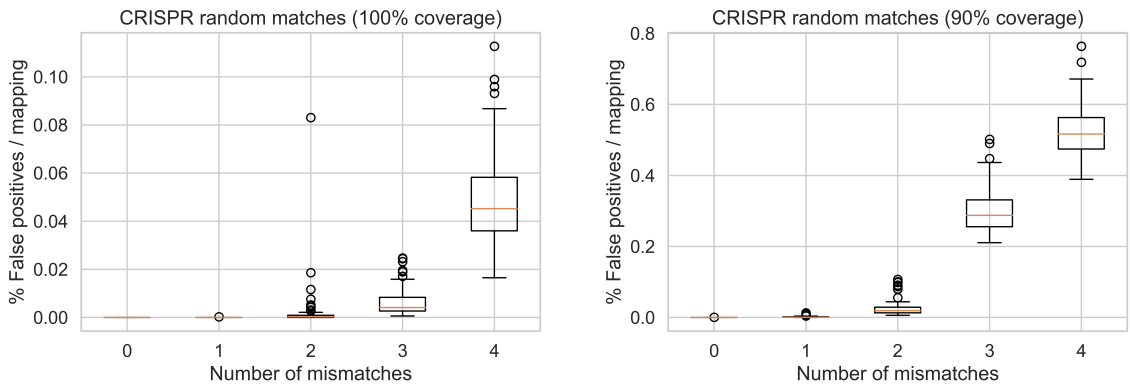
I then measured viral diversity (measured by the number of VCs per isolate) within each phylum (Figure 4.6C). This analysis revealed that the Firmicutes harbour a significantly higher

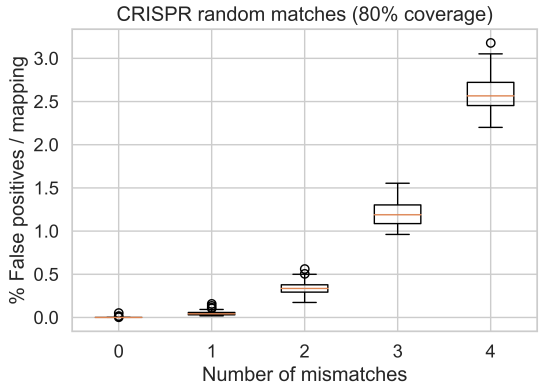
viral diversity, with an average of 3.13 VCs/isolate while also harbouring 60% of the total VCs assigned across all phyla. Interestingly, the Firmicutes diversity was unevenly distributed as most of the viral diversity originated from the Negativicutes and Clostridia classes, with an average of 4.88 VCs and 3.9 VCs per isolate in contrast with the Bacilli (0.99 VCs/isolate), and none for *Bacilli*_A and Desulfitobacteriia classes.

Analysis at the bacterial genus level across all phyla revealed that *Lachnospira*, *Roseburia*, *Agathobacte*r, *Prevotella*, and *Blautia*_A host the highest number of VCs/isolate (Figure 4.6D). With the exception of *Prevotella*, which belongs to the Gram-negative Prevotellaceae family, these genera are members of the Gram-positive Lachnospiraceae family of Firmicutes associated with butyrate-producing spore-formers. In contrast, the lowest viral diversity per isolate was detected among *Helicobacter*, and the lactic acid bacteria *Lactobacillus*, *Lactobacillus*_H, *Enterococcus*_D and *Pediococcus.* Thus, I observe a wide distribution of phage abundance and prevalence across human gut bacteria, even within the same phylum.

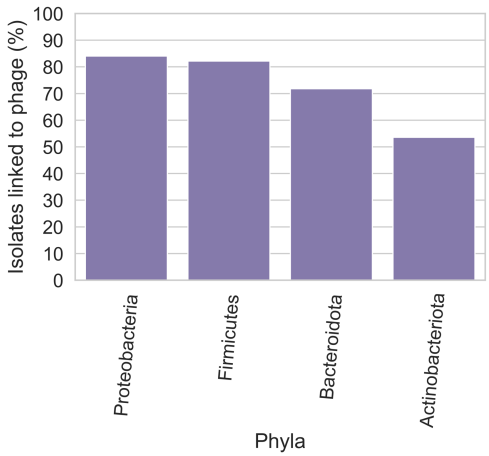
CRISPR spacers can be used to link phages with their host but a limitation is that some bacteria do not encode them and thus their phages will not be detected in the analysis. Although it's estimated that around 46% of bacteria code for CRISPR systems (Karginov and Hannon, 2010), I detected CRISPR spacers in 56.36% of the gut isolate genomes. Despite the discrepancy with the previous estimate, a larger prevalence in the gut may be plausible. It's possible that the incidence of CRISPR systems may vary across different environmental niches.

A

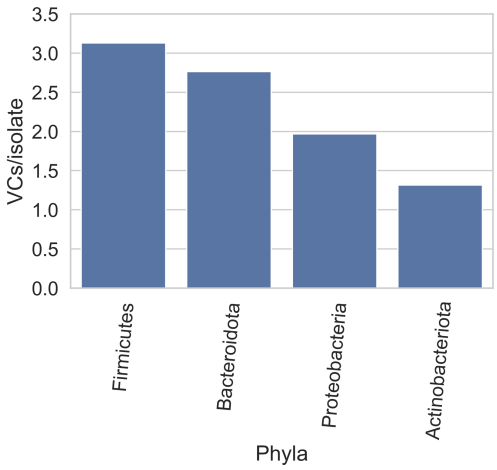



CRISPR random matches (80% coverage)
% False positives / mapping
0.0
0.5
1.0
1.5
2.0
2.5
3.0
0
1
2
3
4
Number of mismatches

B


Isolates linked to phage (%)
0
10
20
30
40
50
60
70
80
90
100
Proteobacteria
Firmicutes
Bacteroidota
Actinobacteriota
Phyla

C


VCs/isolate
0.0
0.5
1.0
1.5
2.0
2.5
3.0
3.5
Firmicutes
Bacteroidota
Proteobacteria
Actinobacteriota
Phyla

D

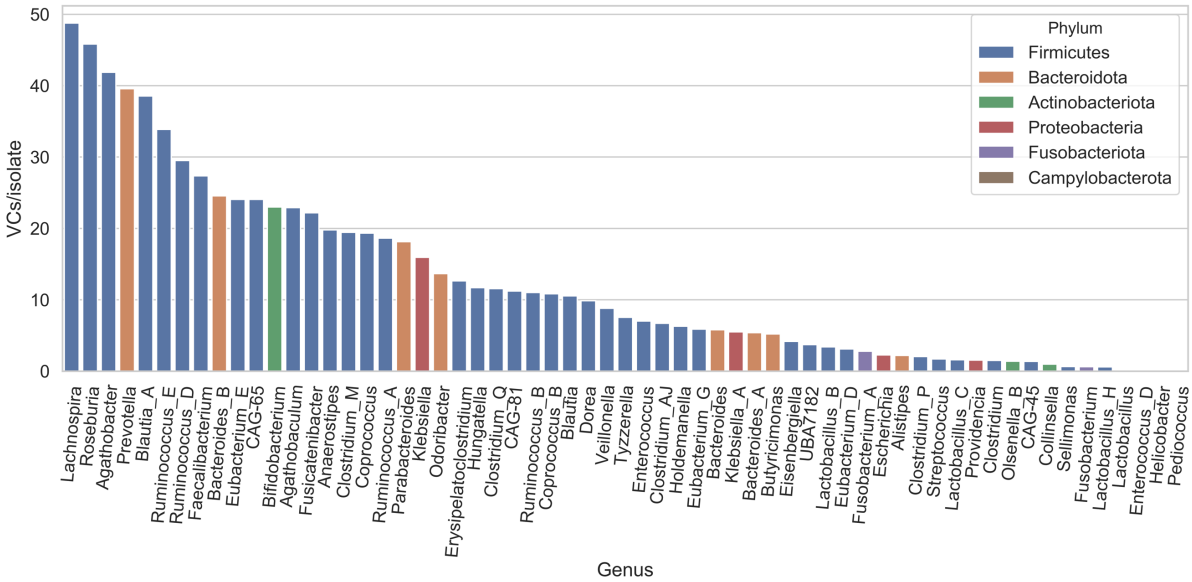

VCs/isolate
0
10
20
30
40
50
Phylum
Firmicutes
Bacteroidota
Actinobacteriota
Proteobacteria
Fusobacteriota
Campylobacterota
Lachnospira
Roseburia
Agathobacter
Prevotella
Blautia_A
Ruminococcus_E
Ruminococcus_D
Faecalibacterium
Bacteroides_B
Eubacterium_E
CAG-65
Bifidobacterium
Agathobaculum
Fusicatenibacter
Anaerostipes
Clostridium_M
Coprococcus
Ruminococcus_A
Parabacteroides
Klebsiella
Odoribacter
Erysipelatoclostridium
Hungatella
Clostridium_Q
CAG-81
Ruminococcus_B
Coprococcus_B
Blautia
Dorea
Veillonella
Tyzzerella
Enterococcus
Clostridium_AJ
Holdemanella
Eubacterium_G
Bacteroides
Klebsiella_A
Bacteroides_A
Butyricimonas
Eisenbergiella
UBA7182
Lactobacillus_B
Eubacterium_D
Fusobacterium_A
Escherichia
Alistipes
Clostridium_P
Streptococcus
Lactobacillus_C
Providencia
Clostridium
Olsenella_B
CAG-45
Collinsella
Sellimonas
Fusobacterium
Lactobacillus_H
Lactobacillus
Enterococcus_D
Helicobacter
Pediococcus
Genus

Figure 4.6. Viral diversity across gut bacteria clades. A) In order to quantify the rate of FPs due to CRISPR random matches, I generated 100 sets of synthetic random spacers and mapped them against the GPD. No FPs were detected at 100% coverage and no mismatches allowed. Across all coverages tested, the rate of FPs increased as more mismatches were allowed. **B)** Percentage of isolates of each phylum linked to phage. Actinobacteriota had the lowest percentage of isolates predicted to be a phage host. Actinobacteriota vs Bacteroidota (P = 0.007, χ^2 test), Actinobacteriota vs Proteobacteria (P = 0.0025, χ^2 test), Actinobacteriota vs Firmicutes (P = 1.01 x 10^{-5}, χ^2 test). **C)** The Firmicutes hosted the highest viral diversity (highest number of VCs/isolate). Firmicutes vs Bacteroidota (P = 0.021, χ^2test), Firmicutes vs Proteobacteria (P = 4.41 x 10^{-6}, χ^2 test), Firmicutes vs Actinobacteriota (P = 1.1 x 10^{-31}, χ^2test). **D)** Bacterial genera with the highest viral diversity were *Lachnospira*, *Roseburia*, *Agathobacter*, *Prevotella*, and *Blautia*_A. On the other hand, the lowest viral diversity was harboured by *Helicobacter* and the lactic acid bacteria *Lactobacillus*, *Lactobacillus*_H, *Enterococcus*_D and *Pediococcus*.

4.2.7 Evaluating host range of gut phages

Horizontal transfer of genes between bacteria via transduction is a major driver of gene flow in bacterial communities (Chen et al., 2018). Host tropism of bacteriophage is believed to be limited by phylogenetic barriers, with most phages being usually restricted to a single host bacterial species (Ackermann, 1998). However, this has not been investigated at large scale across the human gut bacteria. Host assignment at different bacterial taxonomic ranks revealed that the majority of VCs were restricted to infect a single species (64.51%) (Figure 4.7A). I also found many VCs with broader host ranges such as those restricted to a single genus (22.39%), family (10.79%), order (1.86%), class (0.26%) and phylum (0.13%). These findings are in line with a recent survey of the host range of gut phages by meta3C proximity ligation (6,651 unique host-phage pairs) which found that ~69% of gut phages were restricted to a single species (Marbouty et al., 2020). Visualization of very broad range VCs (i.e. those not restricted to a single genus) reveals the large-scale connectivity between phylogenetically distinct bacterial species (Figure 4.7B).

In general, the higher the viral diversity per bacterial genus, the higher the number of phages with broad host range (Spearman's Rho = 0.6685, P= 3.91x10^{-9}) (Figure 4.7C). Even though

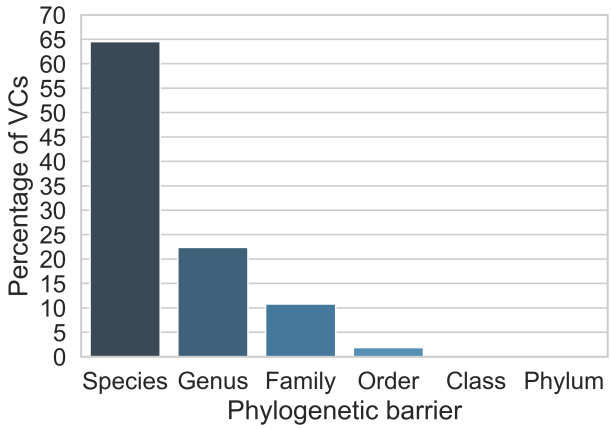
this trend could be explained due to the presence of random matches, as discussed above, no FPs were detected using perfect matches. In addition, when I permuted the labels of the host assignment 300 times, I found the original linear model to significantly deviate from the random one ($P < 0.001$). The average number of broad host range hits for the permuted assignments was 726.9 versus 38.344 for the original assignment, highlighting the containment of phages within bacterial clades.

Surprisingly, two VCs (VC_269 and VC_644) had a host range that spanned two bacterial phyla. VC_269 was predicted to infect *Faecalibacterium prausnitzii*_C (Firmicutes) and two *Bifidobacterium spp.* (Actinobacteriota), while VC_644 had a host range that included 5 *Bacteroides spp*. (Bacteroidota) and *Blautia*_A *wexlerae* (Firmicutes). I predicted VC_269 to be a *Myoviridae* phage, on the other hand, I could not assign a taxonomy rank to VC_644. The presence of integrases in both VCs suggest that these are temperate phages. I hypothesize that additional phages infecting both Actinobacteriota and Firmicutes may be more common, as recent evidence supports a shared ancestry between phages that infect both Actinobacteriota (*Streptomyces*) and Firmicutes (*Faecalibacterium*) (Koert et al., 2019).

Taken together, I reveal that approximately one third of gut phage have a broad host range not limited to a single host species. This analysis provides a comprehensive blueprint of potential phage mediated gene flow networks in human gut microbiome.

The emergence of broad host range phages or 'generalists' has been linked with shifts in bacterial composition linked to nutrient availability (Warwick-Dugdale et al., 2019). In addition, phage generalism has been associated with lower infection efficiency (Howard-Varona et al., 2018). Many members of the gut microbiome are considered copiotrophs based on the copy number of the Ribosomal RNA operon (rrn), as it positively correlates with cellular ribosomal content and maximum growth rate (Gao and Wu, 2018). This would imply that in general, the gut is not a limited nutrient environment and phages can 'secure' a stable host. As stated, the majority of the viral diversity reported here was predicted to infect a single species, which is in line with copiotroph hosts. It's important to consider that some gut bacteria may be oligotrophs as it's increasingly recognized that nutrients in the gut vary spatially (Donaldson et al., 2016). This scenario would probably result in a higher proportion of broad host range for some bacterial species.

A



B

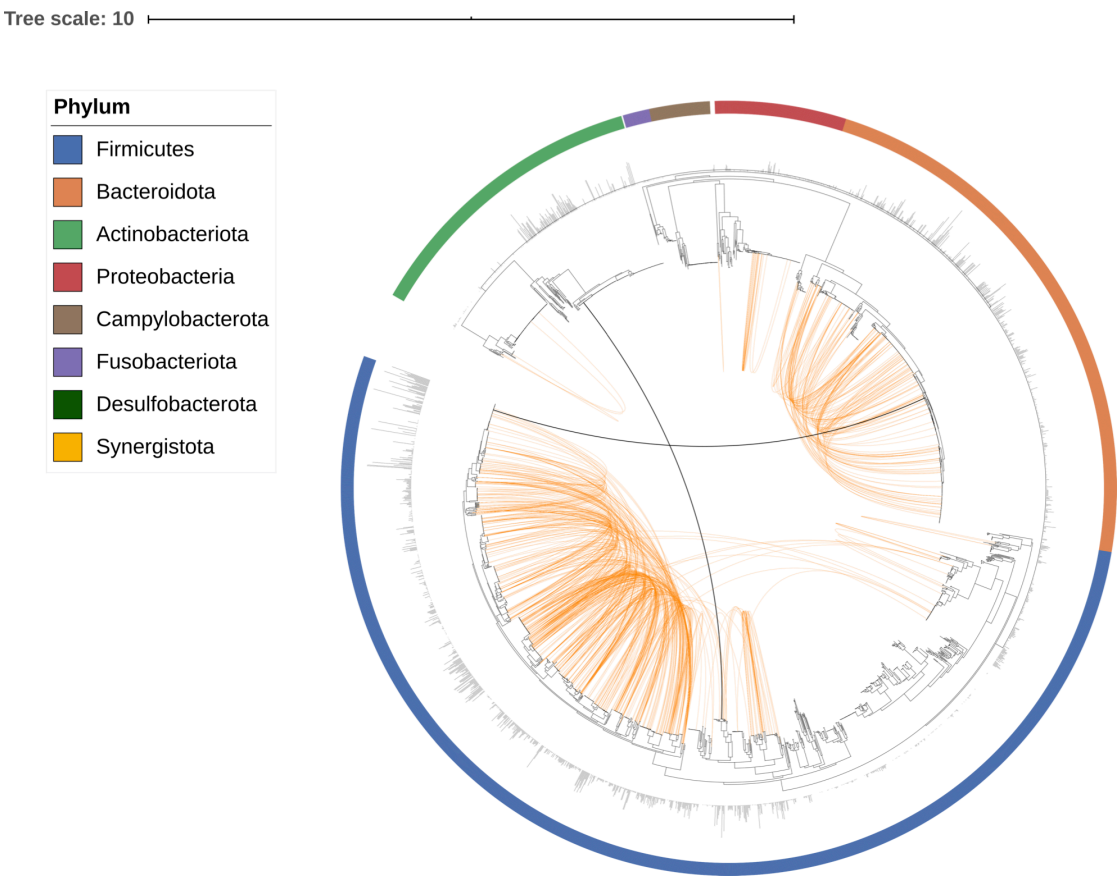


Figure 4.7. Host range of gut phages. A) The majority of VCs were found to be restricted to infect a single species ($P = 0.0$,binomial test). However, a considerable number of VCs (~36%) had a broader host range. **B)** Phylogenetic tree of 2898 gut bacteria isolates showing phage host range. Host assignment was carried out by linking prophages with their assemblies and CRISPR spacer matching. Orange connections represent VCs not restricted to a single genus). Black connections represent VCs able to infect two phyla. Outer bars show phage diversity (VCs/isolate).

C

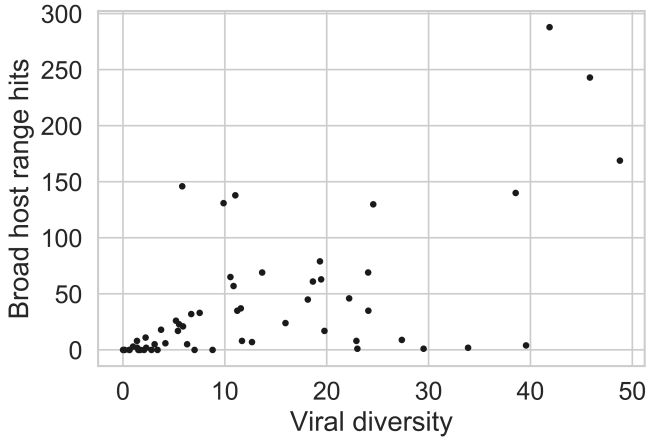


Figure 4.7. Host range of gut phages. C) In general, the higher the viral diversity per bacterial genus, the higher the number of phages with broad host range (Spearman's Rho = 0.6685, P= 3.91x10-9). This trend was significantly different than the one generated from permuting the host assignment labels ($P < 0.001$).

4.3 Conclusions

In this chapter, I carried out a large-scale analysis of gut phages to shed light into their encoded functions. Top viral functions were primarily involved in basic functions of the life cycle of phages such as replication, virion assembly, and lytic enzymes. However, a particular interest of mine was to explore the possibility of gut phages carrying non-canonical viral proteins. In that regard, I found several clades of phages encoding enzymes that participate in sulphur and nucleotide metabolism.

I expect that many of these non-classical viral proteins are involved in promoting a successful infection by energy generation (dissimilatory sulfate reduction) or by manipulating the bacterial nucleotide pool to avoid misincorporation of uracil into the genome of DNA phages. I found that gut phages commonly encode reverse transcriptases (RTs) (~13% of VCs) as opposed to RefSeq phages (<1%). These viral RTs may be fulfilling critical roles in gut phages such as generation of sequence diversity in their receptor binding proteins (RBPs) and protecting lysogens from infection by other phages (superinfection immunity). I also discovered other rare instances (<0.5% of VCs) of phages encoding nutrient uptake genes (e.g. taurine, zinc) which may be of benefit to the bacterial host.

A common issue when analysing metagenomics data is the significant number of proteins annotated as 'hypothetical', hindering efforts to carry out comprehensive functional analyses. This problem is further exacerbated with phages, in part due their large genetic diversity and because many functional experiments have been carried out only in a handful of bacteriophage models (e.g, T4, T7, λ phage). For instance, I found a family of hypothetical proteins present in ~8.5% of all VCs. This observation reflected the lack of annotation for even widespread phage proteins. Despite the limitation regarding functional annotation, I explored the possibility of predicting function for hypothetical viral proteins by exploiting hypervariation motifs. This analysis is particularly suitable for the prediction of RBPs in phages given that the binding domain of RBPs is often under selection to overcome mutations in the bacterial receptor. Using this strategy I was able to identify RBP candidates for two of the most genetically diverse phages in GPD (as measured by genomes per VC), namely the p-crAssphage and the Gubaphage. As hypervariation domains are often found in phages, this

analysis provides a powerful way to narrow down gene function in phages when there is enough availability of viral genetic diversity.

In this chapter I also analysed the Gubaphage clade in detail. Despite the lack of sequence similarity of Gubaphage to p-crAssphage, these phages shared other functional features such as large genome size (>80 kb), *Bacteroides* host range, a BACON-containing protein and a circular genome. Given the high variation of the crAss-like family, these features prompted me to investigate if Gubaphage belonged to a current or novel crAss-like genus or if it was a completely novel clade. By compiling a list of genomes representing all the crAssphage genetic diversity and then constructing a tree using terminase large subunit gene, I discovered that the Gubaphage did not fit any of the previous crAssphage clades. Another interesting feature of Gubaphage was the high number of genomes associated to its VC, suggesting its high prevalence in human metagenomes. Indeed, in the next chapter I use more sensitive methods to confirm its high prevalence across human populations. Elucidation of the functional traits of Gubaphage will require its isolation and characterization as this will help to establish a clearer view of its role in the human gut microbiome.

Having investigated a novel clade of gut phages, I decided to explore the possibility of expanding the diversity of a known phage clade, namely the *Picovirinae* subfamily. In order to study the phylogenetic structure of *Picovirinae* gut phages I computed the fraction of shared PCs among them. This analysis uncovered 4 major phage clades. Notably, all RefSeq classified and several unclassified *Picovirinae* phages were assigned to one of the 4 clades. However, two major clades remained composed of only phages found in GPD. The expansion in diversity of the *Picovirinae* subfamily showcases the importance of metagenomics in filling in diversity gaps in phage taxonomy.

Given the technical challenges when culturing gut bacteria, host assignment of gut phages remains largely unexplored. I opted for two strategies namely CRISPR and prophage matching and in order to minimize false positives, I only considered exact matching. This analysis allowed me to explore viral diversity patterns across different bacterial taxonomic groups. For instance, I found that viral diversity was highest in the Firmicutes while at the genus level, *Lachnospira*, *Roseburia*, and *Agathobacter* harboured the highest number of VCs/isolate, whereas *Enterococcus*_D*, Helicobacter* and *Pediococcus* the least. Notably, I considerably increased the number of phages assigned to less studied bacterial clades. For instance, a search

on “NCBI virus” of phages infecting *Lachnospiraceae* bacteria returns only 8 hits. On the other hand, on this thesis I predicted 2,985 VCs that infected *Lachnospiraceae* bacteria (with an estimated median phage genome completeness of 81.62%).

Although the majority of VCs were found to be restricted to a single bacterial species, a significant percentage (~36%) was predicted to infect multiple species, genera, families, orders, and even classes. A consequence of broad host range phages is an increased connectivity for horizontal gene transfer events between gut bacteria. Since phages can carry genes from their hosts by transduction, broad host range phages can play critical roles in “gene spillage” across very different bacterial clades from the gut microbiome. For instance, a phage can transduce genes from a different family into another bacterial clade. In another transduction event, narrow host range phages (which are more common), can help to move the newly acquired gene into the clade. These events can have important roles in bacterial adaptation in the human gut.

Chapter 5: Global distribution and epidemiology of gut phages

5.1 Introduction and aims

Much of human microbiome research across populations has focused on gut bacteria. Samples from different countries (mainly Western ones), have been analysed for differences in bacterial composition related to health and disease states. In addition, patterns of bacterial profiles have been linked to different factors such as antibiotic use, urbanization, and age. However, epidemiology research of gut phages has been limited and carried out in small cohorts with narrow geographical distribution of samples. Findings to date, include the association of the gut phageome with health and disease, as well as the suggestion of a set of phages carried by at least half of the human population (core virome) (Manrique et al., 2016).

Regarding individual phage clades, efforts have been mainly directed to the analysis of the abundant crAss-like family. For instance, one of the largest studies that analysed the global distribution of crAssphage strains found strong correlations with different clades of gut bacteria, weak associations with diet, but no significant association with health and disease (Edwards et al., 2019).

In this chapter, I analyse global patterns of the human gut phageome and its association with lifestyle and bacterial composition. I then focus on specific VCs, such as those that are widespread across human populations (global) and those that are highly prevalent in individual continents. Finally, I explore the concept of the controversial idea of a core virome using my dataset.

The aims of the research presented in this chapter are:

- assess global patterns of the human gut phageome;
- analyse geographical distribution of relevant VCs;
- assess the concept of a core virome.

5.2 Results and discussion

5.2.1 Saturation curves for VCs

Before proceeding with the analysis of global gut phageome patterns, it was important to assess how much of total viral diversity was captured by GPD predictions (Figure 5.1). With that end, I calculated the number of novel VCs accumulated with the addition of every new sample. By analysing the growth rate of the resultant curve it's possible to estimate the degree of diversity saturation. At the worldwide scale, it seems that GPD reached saturation regarding novel phage diversity. However, this pattern mostly reflects Western continents (64.2% of the samples). When I stratified by continent, in line with the previous finding, Europe and North America seemed to have plateaued. In addition, Asia's and Oceania's curves also showed signs of diversity saturation. In the case of Africa and South America, the diversity appeared to be growing in a linear fashion with each new additional sample, indicating a low degree of saturation. The latter result was expected as the gut phageome of both continents was estimated from only ~200 samples each as opposed to the other continents with thousands of samples. Thus, GPD captured better phage diversity in North America, Europe, Asia and Oceania, while the gut phageome from generally understudied continents such as Africa and South America still remains to be further explored. Importantly, small phages with a genome size < 10 kb (e.g. *Microviridae*) and RNA phages need to be considered for all continents in order to have a fuller picture of the diversity of the gut phageome.

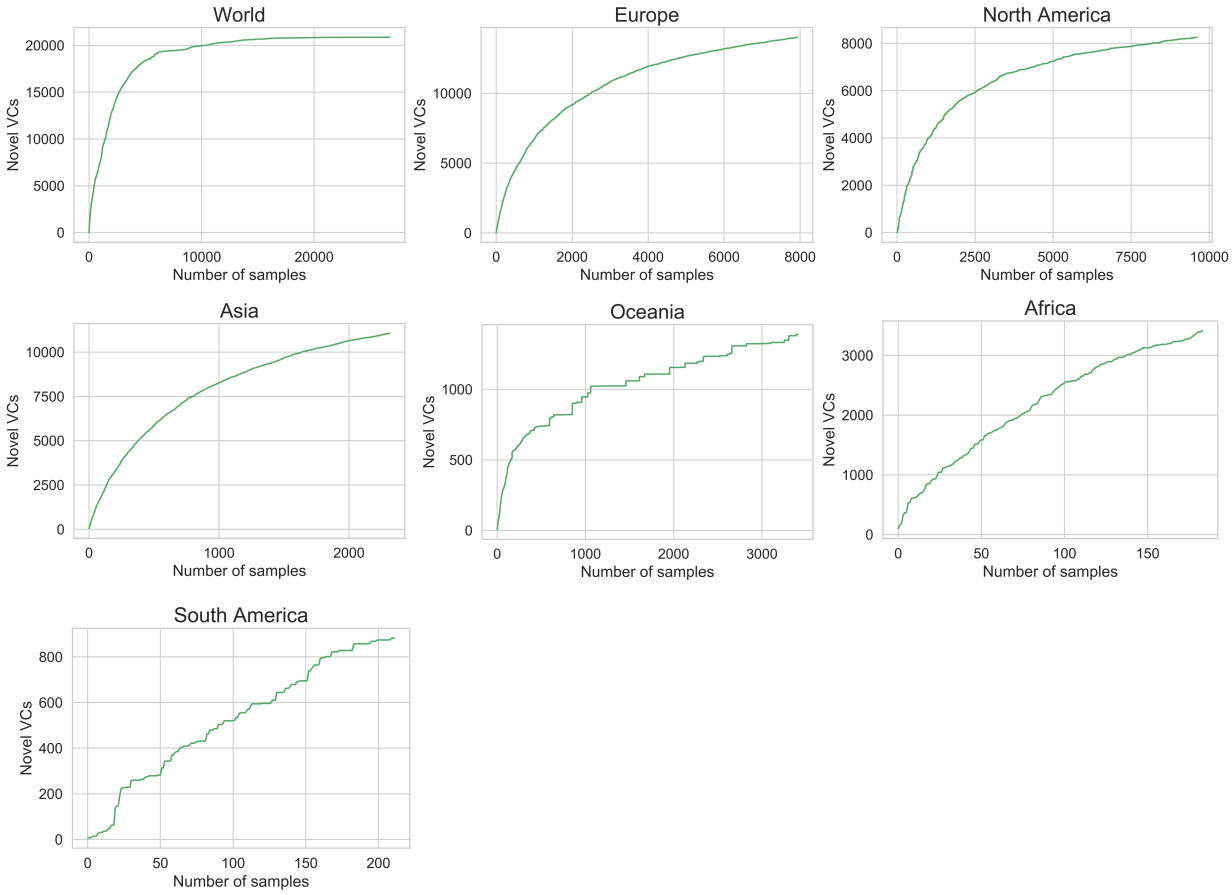


Figure 5.1. Rarefaction curves for viral richness. Saturation curve for viral richness captured in GPD. At the worldwide scale, viral richness seems to have plateaued. However, analysis of individual continents show phage diversity in Africa, and South America is still growing.

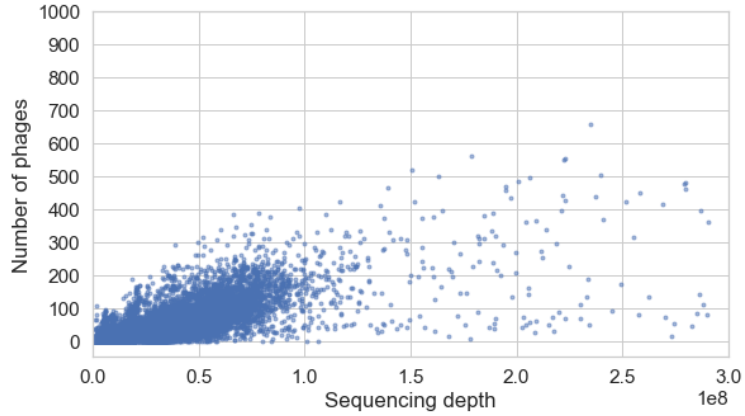
5.2.2 Human lifestyle associated with global gut distribution of phageome types

Each human harbours diverse populations of gut phage, referred to as a phageome. The 28,060 metagenomic datasets used to generate the GPD were sampled from 28 different countries across the six major continents (Africa, Asia, Europe, North America, South America and Oceania) providing a basis to explore patterns in gut phageomes across human populations. I removed samples with a sequencing depth below 50 million reads/sample, as below this threshold I observed a positive correlation between sample depth and number of viral genomes detected (Figure 5.2A). This new subset consisted of 3011 samples and spanned all the continents and 23 countries. I estimated the similarity between samples by computing the number of shared VCs and normalizing it by the total number of VCs in both samples (Jaccard index).

I observed that North American, European, and Asian samples segregated from African and South American samples (Figure 5.2B). Interestingly, this pattern is associated with important differences in human lifestyles. Country-wise, samples derived from Africa and South America come mainly from Peru, Tanzania, and Madagascar. Specifically, Peruvian and Tanzanian samples originate from hunter gatherer communities whereas Malagasy samples come from rural communities with non-Western lifestyles. Oceania was a special case because it had a similar fraction of samples belonging to both groups. However, when I stratified by country, all Fijian samples went to the rural group, whereas Australian samples segregated with the urbanized cluster. Fiji samples were derived from rural agrarian communities. These observations support the hypothesis that lifestyle, particularly urbanization, may drive differences in the gut phageome across different human populations.

I reasoned that the bacterial composition of an individual's microbiome would shape the gut phageome. Prevotellaceae bacteria are more abundant and prevalent in individuals living a rural/traditional lifestyle, whereas *Bacteroides* are more abundant and prevalent in individuals living a urban/Western lifestyle (Wu et al., 2011). By harnessing the host assignment data for each phage, I found that the proportion of VCs assigned to the Prevotellaceae family from African, South American and Fijian samples was much higher than that of North America, Europe, Asia, and Australia (Figure 5.2C). I observed an inverse relationship with *Bacteroides* phage, which were significantly more prevalent in North America, Europe, Asia, and Australia gut microbiomes. Given the correlation of enterotypes and phageome types, driven by the intimate connection between phages and their bacterial hosts, I provide evidence that human lifestyle drives global patterns of gut phageomes by mediating changes in the bacterial gut microbiome.

A



B

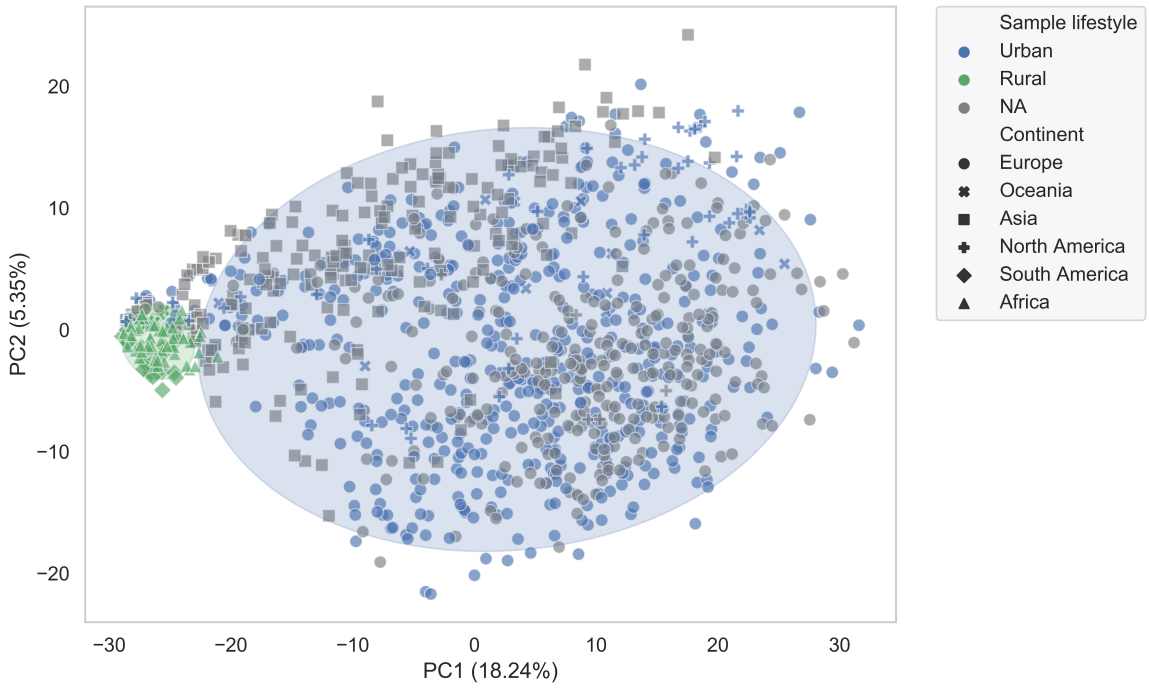


Figure 5.2. Human lifestyle is associated with global gut distribution of phageome types. **A)** Samples exhibit a positive correlation between sequencing depth and number of phage genomes detected. Correlation of samples with sequencing depth < 50 million (Pearson's r: 0.6825, *P* = 0.0). Correlation of samples with sequencing depth > 50 million (Pearson's r: 0.3681, *P* = 2.79e-97). **B)** PCA plot of inter-sample Jaccard distance. Lifestyle is associated with differences in the gut phageome across human populations. Samples from Peru, Madagascar, Tanzania and Fiji are found in the rural cluster whereas those samples with a more Westernized lifestyle (mainly from North America, Europe, and Asia) are found in the urban cluster (P=0.001, $R^2 = 0.36$, PERMANOVA test). Ellipses enclose samples within 2 standard deviations for each lifestyle.

C

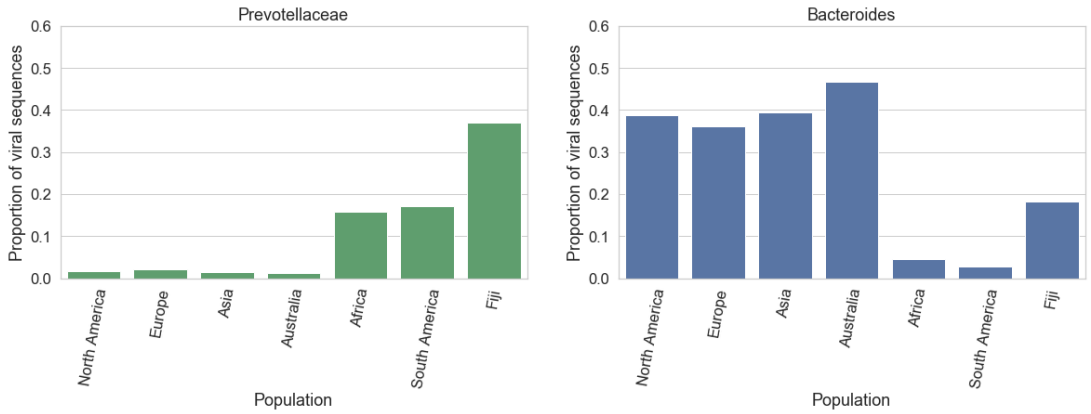


Figure 5.2. Human lifestyle is associated with global gut distribution of phageome types. C) The proportion of VCs that match Prevotellaceae hosts in traditional societies is higher than that of industrialized populations. Conversely, *Bacteroides* hosts are more common in industrialized populations than in traditional societies. Taken together, this result suggests that the composition of the gut phageome at a global scale is driven by the bacterial composition.

5.2.3 Phage carriage across continents

Next, I sought to determine differences in phage carriage according to geographic location (Figure 5.3). It was interesting that despite the large viral diversity that the gut can harbour (21,012 VCs), I detected fewer than 150 VCs in most samples. This threshold could be a result of niche saturation that might prevent exogenous phages from establishing in the gut, mirroring the colonization resistance effect seen in the bacterial gut microbiome. Indeed, longitudinal studies have shown that the gut virome is very stable within individuals (Shkoporov et al., 2019). I did not find significant differences in phage richness across continents except in Africa which had significantly higher diversity.

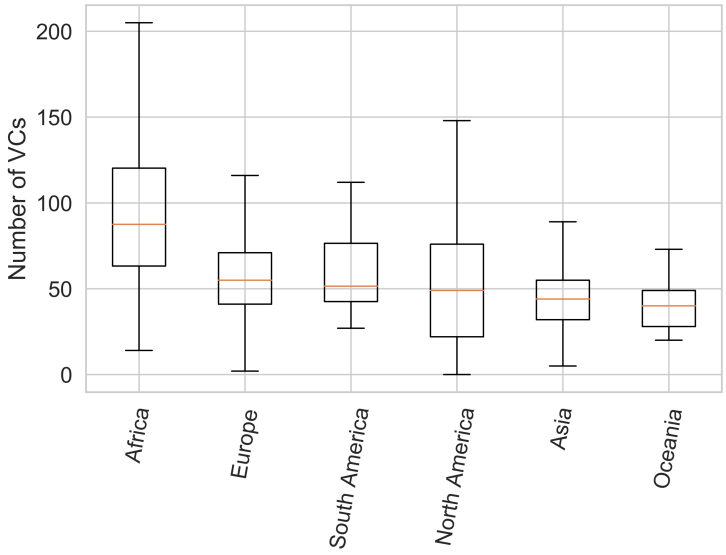


Figure 5.3. Phage carriage across continents. Intra-sample diversity is relatively low compared to the total gut phage diversity. Phage carriage is similar on average per sample across continents except for Africa which is significantly higher. Africa vs Europe (P = 1.82 x 10^{-12}, Mann-Whitney U test), Africa vs South America (P = 0.00033, Mann-Whitney U test), Africa vs North America (P = 4.04 x 10^{-14}, Mann-Whitney U test), Africa vs Asia (P = 6.06 x 10^{-22}, Mann-Whitney U test), Africa vs Oceania (P = 1.64 x 10^{-10}, Mann-Whitney U test)

5.2.4 Uncovering most prevalent phage in global human populations

Stratifying by continent provided me with an unprecedent opportunity to uncover the most prevalent phages around the world. In the case of North America, Europe, and Asia, the host range of the top VCs was dominated by the genera *Bacteroides*, *Bacteroides*_B, and *Parabacteroides*. Notably the p-crAssphage (VC_1) was part of the top VCs for all these continents. Since the gut microbiota of Western societies is dominated by *Bacteroides*, it makes sense that the bacterial hosts of many prevalent VCs are genetically related to this genus. In the case of Africa, South America, and Oceania, for the majority of VCs the bacterial host could not be predicted with the exception of *Faecalibacterium* and *Prevotella*. The absence of host prediction for these continents, may be a consequence of uncultured gut bacteria from these understudied regions, thus hindering efforts to use CRISPR spacers matching or prophage assembly linkage. In general, prevalence of individual VCs was ~25%, the higher prevalence found in South America (~41%) and Oceania (32%) could be result of the limited number of samples to calculate them (<35). Phage prevalence is also dependent on the taxonomic level at

which it's being studied. VCs correspond to subgenus level, however when phages are grouped at genus or family levels their prevalence could substantially increase.

A general observation is that for all continents, phage prevalence follows a power law (Figure 5.4). That is, it appears that across all human populations, there are a few phage clades that are widespread, and they are followed by other clades with decreasing prevalence. Since the rate at which prevalence decreases is proportional to the rank, this behaviour gives rise to a long tail of rare phage clades. High phage prevalence such as that of crAssphage, can be explained by a high prevalence of its bacterial host, while rare phages could be result of them preying on uncommon gut bacteria.

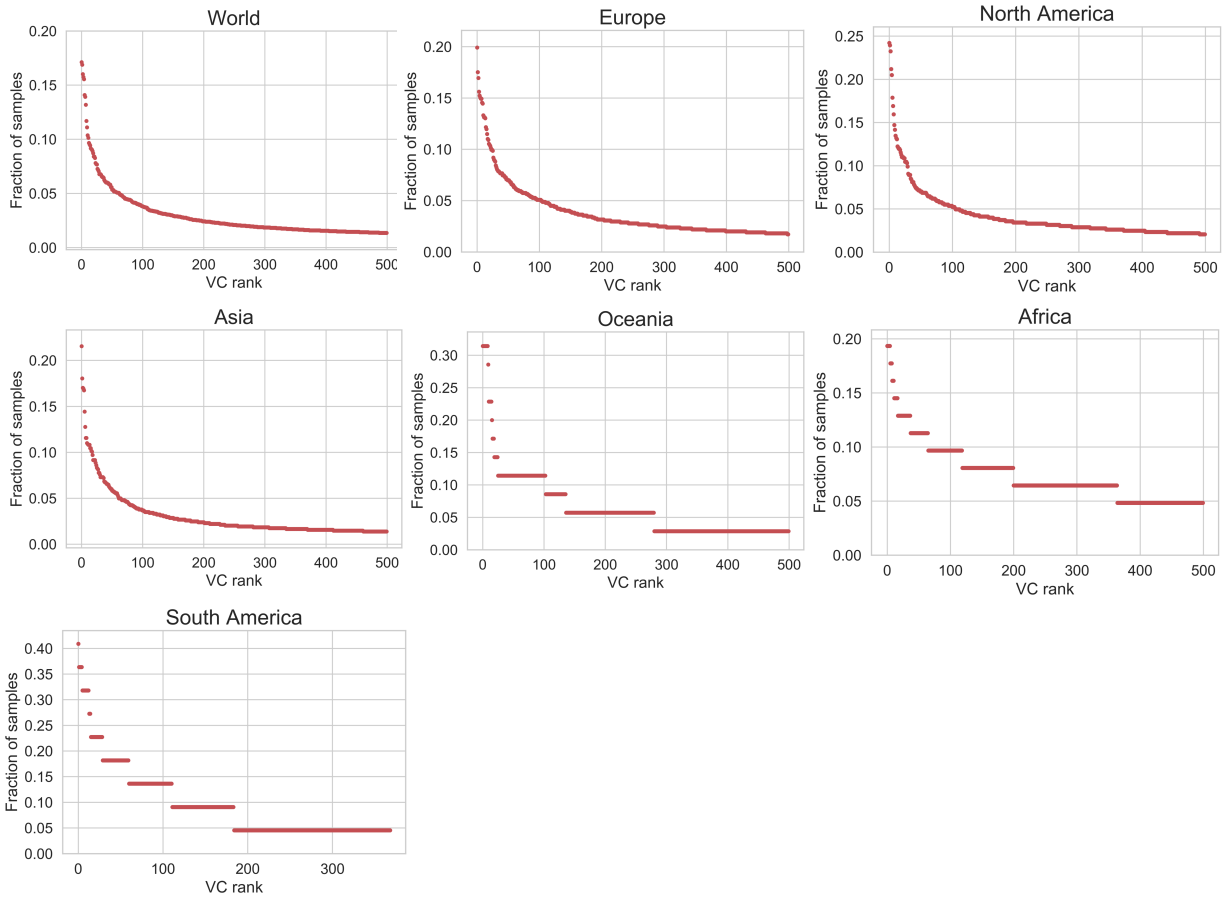


Figure 5.4. Rank prevalence curve for VCs. Prevalence for individual VCs follows a power law distribution across all continents. Phages are usually not found infecting more than ~25% of samples from a given region.

5.2.5 Global distribution of 280 dominant human gut phages

If the gut phageome is predominantly shaped by the bacterial composition, we would expect to observe strong correlation between the prevalence of VCs with that of their bacterial hosts. A clear example is the crAss-like family of gut phages which can be divided into 10 phage genera (Guerin et al., 2018). Genus I, which has been found in a large fraction of Western microbiome samples is able to infect species from the *Bacteroides* genus. In contrast, genera VI, VIII and IX were previously found to be the most prevalent crAss-like phage among Malawian samples (Guerin et al., 2018). Here, I predict that the most probable host of these three phage genera is *Prevotella copri* (rest of crAss-like family predicted hosts in Table 1). In accordance with the results from the Malawian samples, I also found the prevalence of genera VI, VIII and IX to be higher than genus I in Africa and South America (Figure 5.5A). Thus, the crAss-like family is globally distributed with distinct global distribution patterns at the genera level, which appears to be strongly influenced by human lifestyles and enterotypes.

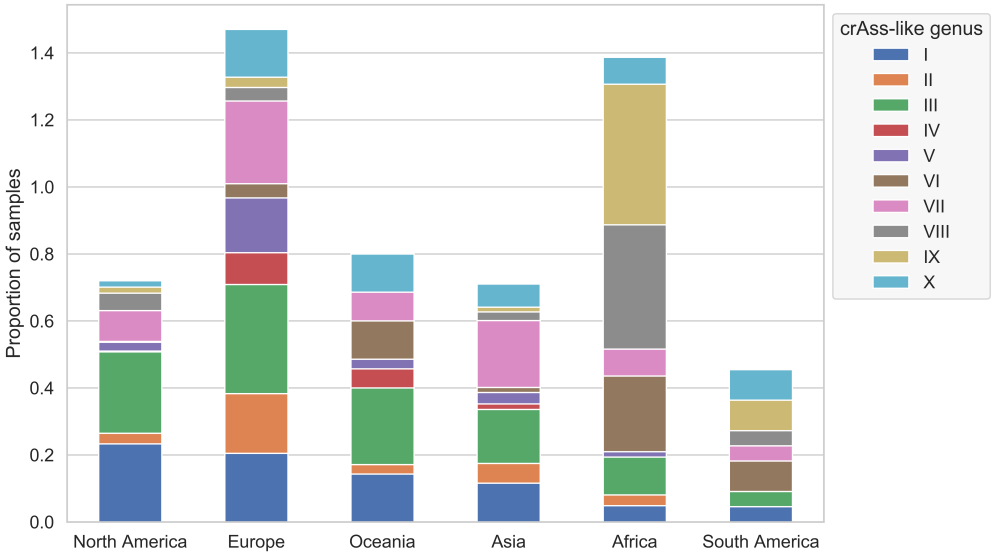
I further investigated if I could identify other gut phage VCs with global distributions. By extending the analysis to all the VCs I was able to detect a total of 280 VCs that were globally distributed (found in at least 5 continents). This represents ~1.3% of all defined VCs (280/21,012). For 119 out of the 280 VCs (42.5%), I was able to classify them to the *Caudovirales* order, whereas the remaining 57.5% remained unclassified. Thus, the majority of globally distributed VCs are completely novel. When I looked at viral families detected within the *Caudovirales*, I detected *Podoviridae* (10 VCs), *Myoviridae* (28 VCs), *Siphoviridae* (43 VCs), and the newly formed family *Herelleviridae* (1 VC). In addition, when I examined at the phage subfamily level, the most common hits corresponded to the *Picovirinae* and *Peduovirinae* subfamilies with 4 VCs each. Importantly, the genomes of 131 members of 57 globally distributed VCs were mined directly from genomes of cultured isolates, providing unique opportunities for follow-up experiments in the lab.

A bacteria-phage network of globally distributed VCs (Figure 5.5B) revealed that *Prevotella* was the most targeted genus (37 VCs), followed by *Faecalibacterium* and *Roseburia* with 15 VCs each. In addition, I observed that in contrast to the Bacteroidales and Oscillospirales, the global VCs associated to the Lachnospirales were highly shared between different genera (Figure 5.5C). Notably, whilst 12 globally distributed VCs were members of the crAss-like family (in black), I was only able to assign a host to 6 VCs which targeted Bacteroidales

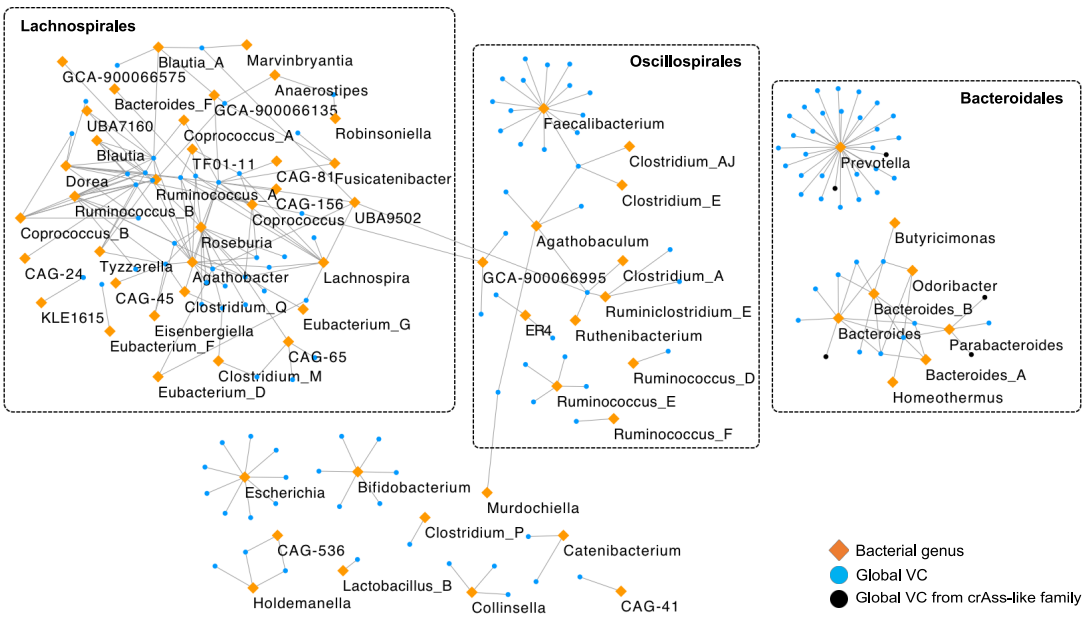
bacteria. I observed that globally distributed phages had a significant broader range (across different genera) than phages found in single continents (P = 1.62x10^{-5}) (Figure 5.5D). This result suggests that broad host-range of certain VCs likely contribute to their expansion across human populations.

Thus, I show that along with 12 crAss-like VCs, there exists a set of at least 280 VCs which are globally distributed. Functional characterization of members of this set will prove useful to shed light on what makes a gut phage to become widespread across human populations.

A



B



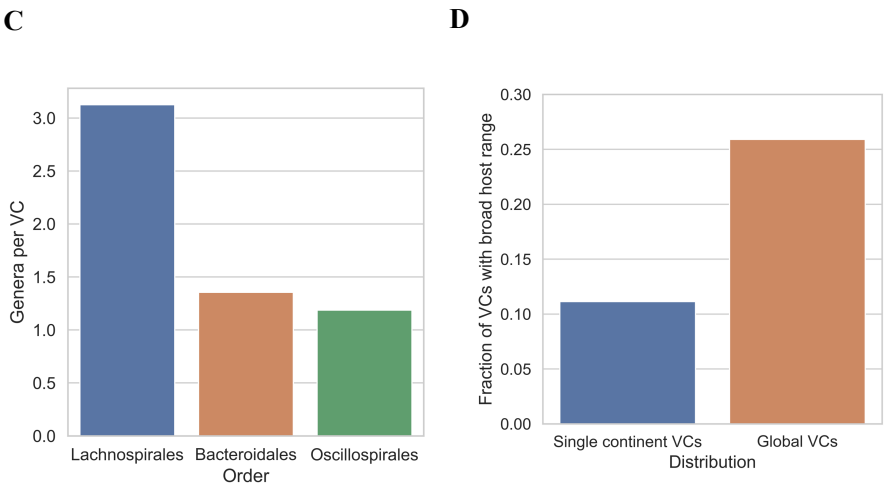


Figure 5.5. Global gut phage clades and their bacterial hosts. A) The crAss-like family is a globally distributed phage. Genera VI, VIII and IX which are predicted to infect a *Prevotella* host are more common in Africa and South America in contrast to genus I which infects a *Bacteroides* host. **B)** Host-phage network of globally distributed VCs (orange) reveals that *Prevotella*, *Faecalibacterium*, and *Roseburia* are the most targeted bacterial genera. VCs that belong to the crAss-like family are highlighted in black; These were predicted to infect *Prevotella*, *Bacteroides*, and *Parabacteroides*. **C)** In contrast to the Bacteroidales and Oscillospirales, the VCs from the Lachnospirales are highly shared. Lachnospirales vs Bacteroidales (P = 9.99 x 10^{-6}, χ^2 test). Lachnospirales vs Oscillospirales (P = 6.55 x 10^{-6}, χ^2 test). **D)** Globally distributed phages had a significantly broader range (above genus) than phages found in single continents (P = 1.63 x 10^{-5}, χ^2 test).

5.2.6 Investigating the concept of a core-virome

Marinque et al. proposed that despite the high interpersonal variation found in the human gut phageome there exists a set of shared phages across individuals (>50%) referred to as the core phageome (Manrique et al., 2016). It was hypothesized that the core phageome is composed of a set of phages which play an important role in maintaining gut microbiome structure/function and thus contribute significantly to human health.

As I showed in Figure 5.4, none of the VCs reached a prevalence >50%, precluding the idea of a core phageome in this work. Nonetheless, I wondered if I could find a reduced set of VCs

that could cover the majority of samples (Figure 5.6A). That is, a sample would be considered covered if at least 1 VC from this set was detected in it. What I found is that at the worldwide level at least one out of 150 VCs were already found in more than 90% of all the samples, and at only 50 VCs the fraction of covered samples was >80% causing the curve to start to plateau. Stratification by continent revealed similar saturation kinetics. At least one out of 50 VCs were found in >50% of samples with the exception of South America (~40%). The more flattened curve observed in South America could be due to the smaller phage genetic diversity captured by GPD. An explanation of why this reduced set of VCs exists is that common phages in the human gut should prey on prevalent bacteria. Certainly, host range prediction of the top 50 VCs for which at least 1 VC is found in >50% of worldwide samples, reveals that these phages infect mostly genera from *Bacteroides*, *Roseburia*, *Parabacteroides*, *Bacteroides_B*, and *Coprococcus*.

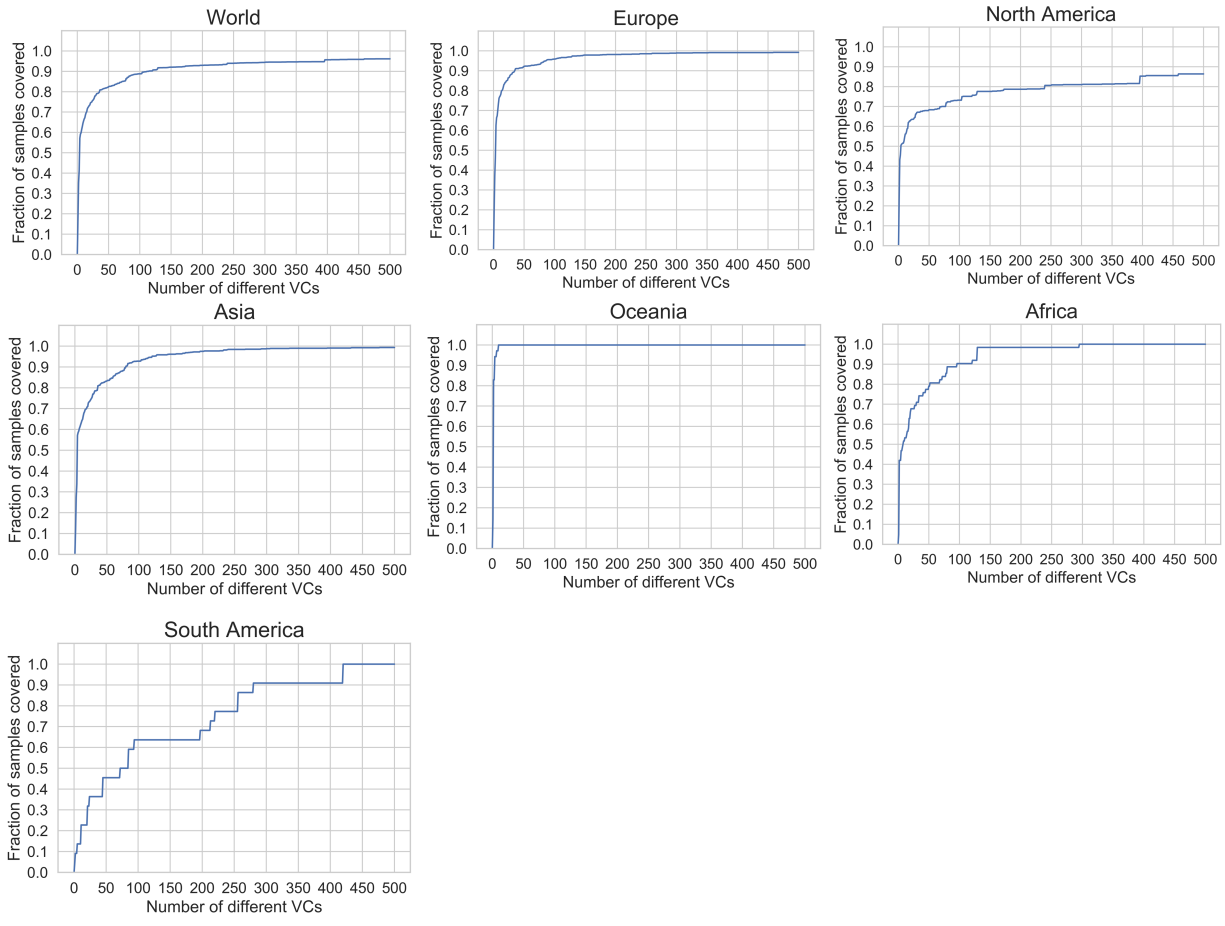
It's also important to mention that although a core virome is unlikely to exist at the ~genus viral level, this finding doesn't reject the idea of highly prevalent viral clades at higher taxonomic ranks. I investigated this idea by measuring the prevalence of the crAss-like family, Gubaphage clade, and *Picovirinae* subfamily across different continents. As we can see in Figure 5.6B, when I pool all the 10 different crAss-like genera, prevalence surpasses ~30% across all continents except in South America, and notably Europe and Africa reach ~70% prevalence. On the other hand, the Gubaphage clade is found well below 20% prevalence across continents, and absent in South America. Europe is the exception with ~40% of samples harbouring a Gubaphage. Finally, I detected the *Picovirinae* subfamily in at least 50% of all samples. Thus, the *Picovirinae* subfamily can be considered a core human phage clade. Notably, its prevalence reaches ~80% in Europe, Africa, and South America. The high prevalence of *Picovirinae* in the last two continents is particularly interesting given that the gut microbiome from African and South American individuals is largely understudied, and thus this finding represents a step forward in understanding and identifying important phages that inhabit their gut.

Analogous to the previous analysis in which I calculated the cumulative fraction of samples covered by each new additional VC, Figure 5.6C shows the same exercise with the crAss-like family, the Gubaphage clade, and the *Picovirinae* subfamily. Combination of the crAss-like family with the Gubaphage clade essentially leaves unchanged the fraction of samples covered when only the crAss-like family is considered, indicating a high co-ocurrence. On the other

hand, when the crAss-like family is combined with the *Picovirinae* subfamily, prevalence surpasses 60% for all continents except in North America (~55%). Notably, Europe and South America reach ~85% prevalence, while in Africa 90% of samples are covered. Combination of the 3 phage clades, does not change much the fraction of samples covered due to the crAss-like and Gubaphage correlation.

Despite only finding one instance of a human core phage (*Picovirinae*), or two if we consider >30% prevalence (*Picovirinae* and crAss-like family), I believe that a proper core phageome may exist. The reason why many studies fail to detect it is because they dereplicate at 95% nucleotide identity. This dereplication threshold is too stringent and thus gives rise to an extremely large variability of the gut phageome (Figure 5.6D). If dereplication was carried out at the level of shared protein clusters (PCs) (e.g. >20% shared PCs), then phage genomes could be clustered at higher phylogenetic levels (genus or subfamily) and phage variation could start to stabilize. Conversely, clustering genomes at very high phylogenetic levels (e.g. order) could result in an unspecific signal.

A



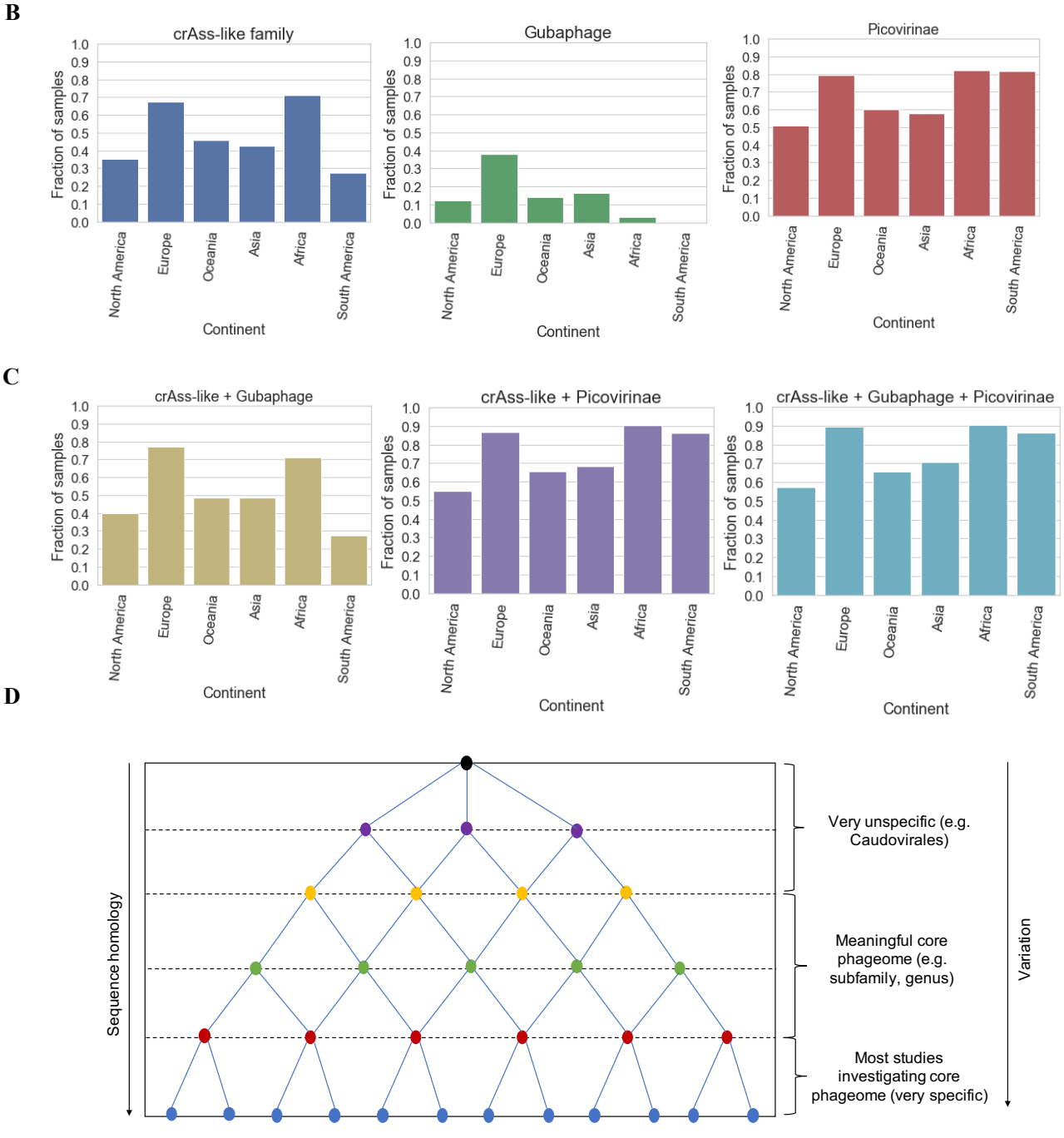


Figure 5.6. Investigating the concept of a core-virome. A) A limited number of VCs are found at least once in a large fraction of human samples across continents. **B)** Analysis of prevalence at higher taxonomic phage clades. CrAss-like phages are found in >30% of worldwide samples, whereas the *Picovirinae* subfamily is found in >50% of samples. **C)** Prevalence analysis with different combinations of the crAss-like family, *Picovirinae*, and Gubaphage clade. **D)** A core phageome may exist, however studies use very stringent dereplication (e.g. 95% nucleotide identity). Probing for higher taxonomic groups may reveal more conserved phages across individuals.

5.3 Conclusions

In this chapter, I analysed the worldwide prevalence and epidemiology of human gut phages by read mapping GPD predictions to a global dataset of human gut metagenomes. This dataset consisted of 3011 samples and spanned all six major continents (Africa, Asia, Europe, North America, South America and Oceania) and 23 countries. The original number of metagenomes considered for this analysis was much bigger (28,060), however samples with a sequencing depth below 50 million reads/sample were removed, as below this threshold I observed a positive correlation between sample depth and number of viral genomes detected. This should be an important consideration for future metagenomic studies of the gut phageome.

I began by studying global patterns of the human gut phageome. A key finding was that urbanization is associated with the composition of the gut phageome. Specifically, when I visualized the distribution of samples, North American, European, and Asian samples segregated from African and South American samples. Samples from the last two continents were derived from communities with non-Western lifestyles. Country-wise stratification showed that Australia belonged to the Western cluster, while Fiji to the rural one. Notably, samples from both countries shared the same lifestyle of their respective cluster. These observations supported the hypothesis that lifestyle, particularly urbanization, may drive differences in the gut phageome across different human populations. In addition, host range prediction of the VCs mapped to each sample, aligned with the expected bacterial enterotype from each continent. Given the correlation of bacterial enterotypes and phageome types, these findings provide evidence that human lifestyle drives global patterns of gut phageomes by mediating changes in the bacterial gut microbiome. Finally, I compared the number of detected VCs per sample across continents. Despite the unprecedented phage diversity found in all samples, I discovered that in general, the majority of individuals only harboured less than 150 VCs.

I then focused on the distribution of individual VCs. A key question was whether there was a set of highly prevalent phage clades which were found across all human populations. For instance, when the p-crAssphage was reported to be found in the majority of analysed samples, a natural question was whether p-crAssphage was a universal highly prevalent phage or if it was exclusive of Western samples. I found that depending on the continent, the most prevalent

phages differed. I found that in North America, Europe, and Asia, p-crAssphage was highly prevalent, but that was not the case for Africa and South America. Nonetheless, for the latter two, I did detect highly prevalent phages that were members of the crAss-like family with a *Prevotella* host range. Despite the dependency of phages on the bacterial composition, I screened for VCs that could be found in all continents. I discovered 280 VCs that were detected in at least 5 continents; a host-phage network showed that the top bacterial genera targeted by these globally distributed VCs were *Prevotella*, *Faecalibacterium*, and *Roseburia*.

The concept of a core virome has sparked controversy in the field, thus I assessed how well it fitted with my data. On one hand, prevalence of individual VCs never reached more than ~25% precluding the idea of a core set of phages shared by at least 50% of individuals. On the other hand, I found that at a worldwide level, at least one of 150 VCs was already found in ~90% of the samples. At the level of continents, at least one of 50 VCs were found in ~50% of the samples. This set of phages is technically not a core virome, but it's surprising the large fraction of samples a relatively small set of VCs can cover given the high level of inter-personal variation found in the gut phageome. A reason why a core virome has not be found may be because analyses are carried out at a very low taxonomic level (e.g. viral species). When I analysed the prevalence of phage clades at a higher taxonomic level, I detected that at least 30% of samples were carrying a crAss-like phage, whereas the *Picovirinae* subfamily was detected in at least 50% of all samples.

Chapter 6: Summary and future work

6.1 Summary

6.1.1 Development of the GPD

In this thesis, I carried out the largest genomic analysis of the human gut phageome by examining more than 142,000 phage genomes derived from 28,060 worldwide distributed human gut metagenomes and 2898 gut bacteria isolates.

In Chapter 3, I introduced the Gut Phage Database (GPD). Although several databases harbouring phage sequences from gut viromes have been published (Gregory et al., 2019; Paez-Espino et al., 2019), to my knowledge, this set represents the largest collection of human gut phage genomes analysed to date. Given the scale of the analyses, not only I was able to identify completely novel viral lineages, but also longer, more complete representatives of known phage genomes. Importantly, this work shows that it is possible to recover high-quality phage genomes from shotgun metagenomes without the need to previously enrich for viral-like particles (VLPs). With this approach, I not only recovered non-integrative phages like *Picovirinae* phages, but also prophage sequences which may rarely enter the lytic cycle and form VLPs. As shotgun metagenomes are far more readily available than VLP metagenomes, I had access to an unparalleled amount of DNA sequences which enabled me to obtain more complete and diverse genomes.

In Chapter 3 I also carried out quality control (QC) and developed methods to handle the massive nature of the dataset. An important finding was the presence of false positives that corresponded to conjugative elements, which highlighted the need for stringent QC when generating thousands of predictions from metagenomic datasets. Even the use of conservative settings of available bioinformatics tools should not preclude the use of extensive QC on phage predictions. As the field moves towards the analysis of larger datasets, manual curation becomes impractical, and I believe that machine learning (ML) approaches (such as the classifier developed here) can be harnessed to help mitigate contamination and significantly boost the quality of the final set of predictions. ML is an extremely fast-paced field and

biologists should take advantage of recent breakthroughs (e.g. deep learning) to make sure that the increasing large volume of biological data submitted to repositories is of high quality (Webb, 2018).

A challenge of this project was the organisation of the large number of predictions into meaningful groups. On one hand, a set of dereplicated predictions at 95% nucleotide identity can be analysed without any further clustering, however patterns can be missed due to underpowering. On the other hand, organising predictions into viral clusters (VCs) allowed me to better generalize my findings. Predictions can be clustered at any defined threshold (e.g. sequence identity), however in order to use a more objective criterion, I benchmarked cluster growth at different thresholds and found that at 90% nucleotide identity most clusters stopped growing (reflecting a more natural threshold). Ideally, clustering by taxonomy proposed by the International Committee on Taxonomy of Viruses (ICTV) should be used (e.g. genus, subfamily), however the majority of my predictions could not be assigned a low level rank or no rank at all. Using a very high-level taxonomy such as order (e.g. *Caudovirales*) also causes to miss patterns because of loss of signal resolution. I expect that as genomic and phenotypical features of the VCs generated are further studied, it's going to be possible to classify them into at least one of the 15 hierarchical ranks recommended by the ICTV (Gorbalenya et al., 2020).

6.1.2 Characterising phage functions and host range

In Chapter 4, I capitalized on the vast number of predictions in GPD to gain knowledge about functions carried out by gut phages. I detected other auxiliary metabolic genes (AMGs) including those involved in nucleotide and sulphur metabolism. Targeted searches also revealed phage reverse transcriptases (RTs) and nutrient transporters.

Mining of function in phages requires a stringent quality control to avoid overestimating their functional potential due to contamination by bacterial genes. Special attention should be paid to genes found at the ends of prophages and contamination assessment should be always carried out. Fortunately, decontamination of phage contigs is becoming automatized with recently published tools such as CheckV (Nayfach et al., 2020) and DRAM-v (Shaffer et al., 2020), facilitating the large-scale annotation of phages from metagenomes. Once a set of clean contigs

are generated, other annotation tools can be used to further characterize the functional potential of phages.

Decontaminated phage contigs still do not guarantee a comprehensive functional annotation as a large fraction of phage proteins are labelled as hypothetical. This limitation highlights our lack of our understanding of protein function which is not exclusive of phages, as recently it was reported that ~27% of proteins derived from gut bacteria do not match any database (Almeida et al., 2020). The number of hypothetical proteins in phages can also be exacerbated by their structural proteins which due to poor conservation are challenging to annotate by conventional methods. However, novel approaches which rely on compositional and physicochemical features such as VIRALpro (Galiez et al., 2016), PVP-SVM (Manavalan et al., 2018), and DeepCapTail (Abid and Zhang, 2018) have showed promise in recognizing them.

The second objective of Chapter 4 was to study relevant gut phage clades. The data-driven discovery of the Gubaphage clade suggests a strategy to identify important clades of phages in metagenomic datasets, as the same approach re-discovered the p-crAssphage as one of the most prevalent clades of human gut phages. Analysis of the *Picovirinae* subfamily illustrated how metagenomics datasets can also help fill-in gaps in viral diversity.

An important element of this work was bacterial host assignment of the majority of gut phages. Both methods used here, exact matches and CRISPR, rely on cultured gut bacteria isolates and highlight the importance of culturing bacteria when studying the viral diversity of ecosystems. The existence of broad host range phages in the human gut suggests that phages have the potential to act as vehicles for horizontal gene transfer (HGT) across distant bacterial clades. The conservative settings used here (100% match and coverage) while highly specific, may have been very stringent and future work could be benefited by allowing a small number of mismatches while maintaining a high specificity.

6.1.3 Epidemiology of gut phages

In Chapter 5, I investigated the epidemiology of gut phages. To my knowledge this is the most comprehensive analysis regarding the global distribution of gut phages given the diversity of the metagenomes (6 continents and 23 countries) and number of phages clades taken into account (21,012 VCs). At a global scale, I provided evidence that the composition of the gut phageome depends on the associated lifestyle of a sample, but also on the gut bacterial composition carried by an individual.

The general dependency of the gut phageome on bacterial composition does not preclude the idea of a global highly prevalent clade of phages (e.g. a VC with a very broad host range). Since its discovery in 2014 (Dutilh et al., 2014), the p-crAssphage has attracted the attention of the microbiome field and even taken as a biomarker of human faecal contamination. After analysing the most prevalent VCs per continent, I discovered that the p-crAssphage was not a highly prevalent clade in Africa and South America. This result provided evidence that p-crAssphage is not a highly prevalent phage in the gut of individuals with a non-Western lifestyle. However, when I analysed the whole crAss-like family, I found some of its members (particularly genera VI,VIII, and IX) in Africa and South America. Host prediction of these phage genera revealed that they prey on *Prevotella copri*. Therefore, it seems that the crAss-like family is a highly prevalent clade of gut phages around the world, raising questions of the biological adaptations that contribute to its success.

This result also highlighted the need to cluster phages into higher taxonomic groups (e.g. genus, subfamily, family) when studying general patterns in the gut phageome. The reason why many studies have not found a core phageome may be because they dereplicate contigs at the species level (e.g. 95% nucleotide identity). This threshold is too stringent; seemingly unrelated phages at the nucleotide sequence level (such as the members of the crAss-like family) may constitute a well-defined clade of phages that share a significant fraction of protein clusters.

When I analysed the concept of a core phageome using VCs, I couldn't find a single VC that was found in more than 50% of samples. However, when I analysed at the phage subfamily level, I found that the *Picovirinae* clade qualified to be a member of the core phageome. Importantly, this clade was found in over 80% of samples from Africa and South America which gut microbiomes are largely unexplored.

6.2 Main findings of this work

1. With proper QC measures, mining of shotgun metagenomes can generate highly complete representative phage genomes complementing VLP enriched metagenomes.
2. A large fraction of gut phages often encode reverse transcriptases (RTs) and auxiliary metabolic genes (AMGs) involved in nucleotide and sulphur metabolism.
3. The Gubaphage clade is a novel gut phage with reminiscent features to crAssphage and is globally distributed.
4. Metagenomics can be harnessed to expand and increase the resolution of previously defined phage subfamilies (*Picovirinae* subfamily).
5. A significant fraction of gut phages (~36%) are not restricted to infect a single species, potentially facilitating gene flow networks between phylogenetically distinct gut bacteria.
6. At a global scale, the gut phageome is associated to lifestyle and influenced by the gut bacterial composition.
7. P-crAssphage is not a highly prevalent phage in Africa and South America, but other members of the crAss-like family that infect *Prevotella copri*.
8. A group of core phages may exist at a global scale (such as the *Picovirinae* subfamily), and may become apparent when dereplicating at higher phage taxonomic ranks.

6.3 Future work

1. *Organizing phage diversity to improve knowledge transfer across metagenomic studies*

   With the current wealth of phage genomes stored in metagenomes, it's now possible to start organizing the large number of phage sequences into meaningful clusters which represent high level candidate viral clades (e.g. subfamilies). This organisation would facilitate the detection of common phage clades across conditions and environments (e.g. is there a phage shared by all body sites?)

2. *Elucidating the extent of active prophages in the human gut*

   An outstanding question is whether prophage sequences integrated in gut bacteria are active or not. Prophages can become "grounded" by mutations in integrases or can accumulate deleterious mutations in essential genes. Conversely, some prophage genes can be useful to bacteria and thus their function is conserved (domestication). Analysis of positive and negative selection on prophage genes from gut bacteria could shed light on this matter.

3. *Mining of phage-encoded antimicrobials*

   Phages represent a rich source of antimicrobials. Given that over 40,000 GPD phage genomes were assigned a host, custom phage encoded antimicrobials such as endolysins can be predicted for hundreds of gut bacteria species. This large-scale resource of anti-bacterial proteins could lead to the development of therapies that specifically modulate the composition of the human gut microbiota.

4. *Investigating diversity of Microviridae/RNA gut phages*

   Due to the minimum genome size imposed in GPD (10 kb), *Microviridae* phages were not investigated in this work. Smaller contigs could be re-analysed and further supported by other tools such as CheckV or an ensemble of predictions tools such as

the What the Phage workflow (Marquet et al., 2020). In the case of RNA gut phages, metatranscriptomics datasets could be harnessed for their discovery.

5. *Wet-lab validation of findings*

This thesis generated a vast amount of predictions that can guide experiments in the laboratory. Since many GPD phages are found in publicly available gut bacteria, further investigation in the wet lab can be carried out on the predicted host range of gut phages and functions conferred by phage-encoded auxiliary metabolic genes.

References

Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., Devin, M., Ghemawat, S., Irving, G., Isard, M., et al. (2016). TensorFlow: a system for large-scale machine learning. In Proceedings of the 12th USENIX Conference on Operating Systems Design and Implementation, (Savannah, GA, USA: USENIX Association), pp. 265–283.

Abedon, S.T., and LeJeune, J.T. (2007). Why Bacteriophage Encode Exotoxins and other Virulence Factors. Evol. Bioinforma. Online 1, 97–110.

Abedon, S.T., Kuhl, S.J., Blasdel, B.G., and Kutter, E.M. (2011). Phage treatment of human infections. Bacteriophage 1, 66–85.

Ackermann, H.W. (1998). Tailed bacteriophages: the order caudovirales. Adv. Virus Res. 51, 135–201.

Adriaenssens, E., and Brister, J.R. (2017). How to Name and Classify Your Phage: An Informal Guide. Viruses 9.

Alawneh, A.M., Qi, D., Yonesaki, T., and Otsuka, Y. (2016). An ADP-ribosyltransferase Alt of bacteriophage T4 negatively regulates the Escherichia coli MazF toxin of a toxin-antitoxin module. Mol. Microbiol. 99, 188–198.

Altamirano, F.L.G., and Barr, J.J. (2019). Phage Therapy in the Postantibiotic Era. Clinical Microbiology Reviews 32.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. (1990). Basic local alignment search tool. J. Mol. Biol. 215, 403–410.

Anantharaman, K., Duhaime, M.B., Breier, J.A., Wendt, K.A., Toner, B.M., and Dick, G.J. (2014). Sulfur oxidation genes in diverse deep-sea viruses. Science 344, 757–760.

Avrani, S., Schwartz, D.A., and Lindell, D. (2012). Virus-host swinging party in the oceans. Mob. Genet. Elem. 2, 88–95.

Bankevich, A., Nurk, S., Antipov, D., Gurevich, A.A., Dvorkin, M., Kulikov, A.S., Lesin, V.M., Nikolenko, S.I., Pham, S., Prjibelski, A.D., et al. (2012). SPAdes: A New Genome Assembly Algorithm and Its Applications to Single-Cell Sequencing. J Comput Biol 19, 455–477.

Barr, J.J., Auro, R., Furlan, M., Whiteson, K.L., Erb, M.L., Pogliano, J., Stotland, A., Wolkowicz, R., Cutting, A.S., Doran, K.S., et al. (2013). Bacteriophage adhering to mucus provide a non-host-derived immunity. Proc. Natl. Acad. Sci. U. S. A. 110, 10771–10776.

Barrangou, R., Fremaux, C., Deveau, H., Richards, M., Boyaval, P., Moineau, S., Romero, D.A., and Horvath, P. (2007). CRISPR provides acquired resistance against viruses in prokaryotes. Science 315, 1709–1712.

Bi, D., Xu, Z., Harrison, E.M., Tai, C., Wei, Y., He, X., Jia, S., Deng, Z., Rajakumar, K., and Ou, H.-Y. (2012). ICEberg: a web-based resource for integrative and conjugative elements found in Bacteria. Nucleic Acids Res 40, D621–D626.

Bin Jang, H., Bolduc, B., Zablocki, O., Kuhn, J.H., Roux, S., Adriaenssens, E.M., Brister, J.R., Kropinski, A.M., Krupovic, M., Lavigne, R., et al. (2019). Taxonomic assignment of uncultivated prokaryotic virus genomes is enabled by gene-sharing networks. Nat. Biotechnol. 37, 632–639.

Blower, T.R., Evans, T.J., Przybilski, R., Fineran, P.C., and Salmond, G.P.C. (2012). Viral Evasion of a Bacterial Suicide System by RNA–Based Molecular Mimicry Enables Infectious Altruism. PLoS Genet. 8.

Bobay, L.-M., Touchon, M., and Rocha, E.P.C. (2014). Pervasive domestication of defective prophages by bacteria. Proc. Natl. Acad. Sci. U. S. A. 111, 12127–12132.

Bondy-Denomy, J., Pawluk, A., Maxwell, K.L., and Davidson, A.R. (2013). Bacteriophage genes that inactivate the CRISPR/Cas bacterial immune system. Nature 493, 429–432.

Bondy-Denomy, J., Qian, J., Westra, E.R., Buckling, A., Guttman, D.S., Davidson, A.R., and Maxwell, K.L. (2016). Prophages mediate defense against phage infection through diverse mechanisms. ISME J. 10, 2854–2866.

Breitbart, M., Thompson, L., Suttle, C., and Sullivan, M. (2007). Exploring the Vast Diversity of Marine Viruses. Oceanography 20, 135–139.

Brister, J.R., Ako-adjei, D., Bao, Y., and Blinkova, O. (2015). NCBI Viral Genomes Resource. Nucleic Acids Res 43, D571–D577.

Browne, H.P., Forster, S.C., Anonye, B.O., Kumar, N., Neville, B.A., Stares, M.D., Goulding, D., and Lawley, T.D. (2016). Culturing of 'unculturable' human microbiota reveals novel taxa and extensive sporulation. Nature 533, 543–546.

Brüssow, H., and Hendrix, R.W. (2002). Phage genomics: small is beautiful. Cell 108, 13–16.

Burmeister, A.R., Fortier, A., Roush, C., Lessing, A.J., Bender, R.G., Barahman, R., Grant, R., Chan, B.K., and Turner, P.E. (2020). Pleiotropy complicates a trade-off between phage resistance and antibiotic resistance. Proc. Natl. Acad. Sci. 117, 11207–11216.

Canchaya, C., Fournous, G., Chibani-Chennoufi, S., Dillmann, M.L., and Brüssow, H. (2003). Phage as agents of lateral gene transfer. Curr. Opin. Microbiol. 6, 417–424.

Chan, C.X., Mahbob, M., and Ragan, M.A. (2013). Clustering evolving proteins into homologous families. BMC Bioinformatics 14, 120.

Chatterjee, A., and Duerkop, B.A. (2019). Sugar and fatty acids Ack-celerate prophage induction. Cell Host Microbe 25, 175–176.

Chaumeil, P.-A., Mussig, A.J., Hugenholtz, P., and Parks, D.H. (2019). GTDB-Tk: a toolkit to classify genomes with the Genome Taxonomy Database. Bioinformatics.

Chen, J., Quiles-Puchalt, N., Chiang, Y.N., Bacigalupe, R., Fillol-Salom, A., Chee, M.S.J., Fitzgerald, J.R., and Penadés, J.R. (2018). Genome hypermobility by lateral transduction. Science 362, 207–212.

Chen, Y., Golding, I., Sawai, S., Guo, L., and Cox, E.C. (2005). Population Fitness and the Regulation of Escherichia coli Genes by Bacterial Viruses. PLoS Biol. 3.

Chopin, M.-C., Chopin, A., and Bidnenko, E. (2005). Phage abortive infection in lactococci: variations on a theme. Curr. Opin. Microbiol. 8, 473–479.

Clokie, M.R.J., and Mann, N.H. (2006). Marine cyanophages and light. Environ. Microbiol. 8, 2074–2082.

Clooney, A.G., Sutton, T.D.S., Shkoporov, A.N., Holohan, R.K., Daly, K.M., O'Regan, O., Ryan, F.J., Draper, L.A., Plevy, S.E., Ross, R.P., et al. (2019). Whole-Virome Analysis Sheds Light on Viral Dark Matter in Inflammatory Bowel Disease. Cell Host Microbe 26, 764-778.e5.

Cornuault, J.K., Petit, M.-A., Mariadassou, M., Benevides, L., Moncaut, E., Langella, P., Sokol, H., and De Paepe, M. (2018). Phages infecting Faecalibacterium prausnitzii belong to novel viral genera that help to decipher intestinal viromes. Microbiome 6, 65.

Couvin, D., Bernheim, A., Toffano-Nioche, C., Touchon, M., Michalik, J., Néron, B., Rocha, E.P.C., Vergnaud, G., Gautheret, D., and Pourcel, C. (2018). CRISPRCasFinder, an update of CRISRFinder, includes a portable version, enhanced performance and integrates search for Cas proteins. Nucleic Acids Res. 46, W246–W251.

D'Herelle, F. (2007). On an invisible microbe antagonistic toward dysenteric bacilli: brief note by Mr. F. D'Herelle, presented by Mr. Roux. 1917. Res. Microbiol. 158, 553–554.

de Jonge, P.A., von Meijenfeldt, F.A.B., van Rooijen, L.E., Brouns, S.J.J., and Dutilh, B.E. (2019). Evolution of BACON Domain Tandem Repeats in crAssphage and Novel Gut Bacteriophage Lineages. Viruses 11, 1085.

De Sordi, L., Khanna, V., and Debarbieux, L. (2017). The Gut Microbiota Facilitates Drifts in the Genetic Diversity and Infectivity of Bacterial Viruses. Cell Host Microbe 22, 801-808.e3.

Delavat, F., Miyazaki, R., Carraro, N., Pradervand, N., and van der Meer, J.R. (2017). The hidden life of integrative and conjugative elements. FEMS Microbiol. Rev. 41, 512–537.

Devoto, A.E., Santini, J.M., Olm, M.R., Anantharaman, K., Munk, P., Tung, J., Archie, E.A., Turnbaugh, P.J., Seed, K.D., Blekhman, R., et al. (2019). Megaphages infect Prevotella and variants are widespread in gut microbiomes. Nat. Microbiol. 4, 693–700.

Diard, M., Bakkeren, E., Cornuault, J.K., Moor, K., Hausmann, A., Sellin, M.E., Loverdo, C., Aertsen, A., Ackermann, M., De Paepe, M., et al. (2017). Inflammation boosts bacteriophage transfer between Salmonella spp. Science 355, 1211–1215.

Dion, M.B., Oechslin, F., and Moineau, S. (2020). Phage diversity, genomics and phylogeny. Nat. Rev. Microbiol. 18, 125–138.

Donaldson, G.P., Lee, S.M., and Mazmanian, S.K. (2016). Gut biogeography of the bacterial microbiota. Nat. Rev. Microbiol. 14, 20–32.

Dongen (S.M.), van (2000). Graph Clustering by Flow Simulation (Universiteit Utrecht).

Dowah, A.S.A., and Clokie, M.R.J. (2018). Review of the nature, diversity and structure of bacteriophage receptor binding proteins that target Gram-positive bacteria. Biophys. Rev. 10, 535–542.

Dunne, M., Rupf, B., Tala, M., Qabrati, X., Ernst, P., Shen, Y., Sumrall, E., Heeb, L., Plückthun, A., Loessner, M.J., et al. (2019). Reprogramming Bacteriophage Host Range through Structure-Guided Design of Chimeric Receptor Binding Proteins. Cell Rep. 29, 1336-1350.e4.

Dutilh, B.E., Cassman, N., McNair, K., Sanchez, S.E., Silva, G.G.Z., Boling, L., Barr, J.J., Speth, D.R., Seguritan, V., Aziz, R.K., et al. (2014). A highly abundant bacteriophage discovered in the unknown sequences of human faecal metagenomes. Nat. Commun. 5, 4498.

Džunková, M., Low, S.J., Daly, J.N., Deng, L., Rinke, C., and Hugenholtz, P. (2019). Defining the human gut host–phage network through single-cell viral tagging. Nat. Microbiol. 4, 2192–2203.

Eddy, S.R. (1998). Profile hidden Markov models. Bioinformatics 14, 755–763.

Edwards, R.A., McNair, K., Faust, K., Raes, J., and Dutilh, B.E. (2016). Computational approaches to predict bacteriophage-host relationships. FEMS Microbiol. Rev. 40, 258–272.

Edwards, R.A., Vega, A.A., Norman, H.M., Ohaeri, M., Levi, K., Dinsdale, E.A., Cinek, O., Aziz, R.K., McNair, K., Barr, J.J., et al. (2019). Global phylogeography and ancient evolution of the widespread human gut virus crAssphage. Nat. Microbiol. 4, 1727–1736.

Enault, F., Briet, A., Bouteille, L., Roux, S., Sullivan, M.B., and Petit, M.-A. (2017). Phages rarely encode antibiotic resistance genes: a cautionary tale for virome analyses. ISME J. 11, 237–247.

Feiner, R., Argov, T., Rabinovich, L., Sigal, N., Borovok, I., and Herskovits, A.A. (2015). A new perspective on lysogeny: prophages as active regulatory switches of bacteria. Nat. Rev. Microbiol. 13, 641–650.

Fernandes, S., and São-José, C. (2018). Enzymes and Mechanisms Employed by Tailed Bacteriophages to Breach the Bacterial Cell Barriers. Viruses 10.

Fokine, A., and Rossmann, M.G. (2014). Molecular architecture of tailed double-stranded DNA phages. Bacteriophage 4.

Fraser, J.S., Yu, Z., Maxwell, K.L., and Davidson, A.R. (2006). Ig-like domains on bacteriophages: a tale of promiscuity and deceit. J. Mol. Biol. 359, 496–507.

Gao, Y., and Wu, M. (2018). Free-living Bacterial Communities Are Mostly Dominated by Oligotrophs. BioRxiv 350348.

Gödeke, J., Paul, K., Lassak, J., and Thormann, K.M. (2011). Phage-induced lysis enhances biofilm formation in Shewanella oneidensis MR-1. ISME J. 5, 613–626.

Gorbalenya, A.E., Krupovic, M., Mushegian, A., Kropinski, A.M., Siddell, S.G., Varsani, A., Adams, M.J., Davison, A.J., Dutilh, B.E., Harrach, B., et al. (2020). The new scope of virus taxonomy: partitioning the virosphere into 15 hierarchical ranks. Nat. Microbiol. 5, 668–674.

Gregory, A.C., Zablocki, O., Howell, A., Bolduc, B., and Sullivan, M.B. (2019). The human gut virome database. BioRxiv 655910.

Guerin, E., Shkoporov, A., Stockdale, S.R., Clooney, A.G., Ryan, F.J., Sutton, T.D.S., Draper, L.A., Gonzalez-Tortuero, E., Ross, R.P., and Hill, C. (2018). Biology and Taxonomy of crAss-like Bacteriophages, the Most Abundant Virus in the Human Gut. Cell Host Microbe 24, 653-664.e6.

Guglielmini, J., de la Cruz, F., and Rocha, E.P.C. (2013). Evolution of Conjugation and Type IV Secretion Systems. Mol. Biol. Evol. 30, 315–331.

Hall, A.R., Scanlan, P.D., Morgan, A.D., and Buckling, A. (2011). Host-parasite coevolutionary arms races give way to fluctuating selection. Ecol. Lett. 14, 635–642.

Hampton, H.G., Watson, B.N.J., and Fineran, P.C. (2020). The arms race between bacteria and their phage foes. Nature 577, 327–336.

Harvey, H., Bondy-Denomy, J., Marquis, H., Sztanko, K.M., Davidson, A.R., and Burrows, L.L. (2018). Pseudomonas aeruginosa defends against phages through type IV pilus glycosylation. Nat. Microbiol. 3, 47–52.

Hizi, A., and Herzig, E. (2015). dUTPase: the frequently overlooked enzyme encoded by many retroviruses. Retrovirology 12, 70.

Howard-Varona, C., Hargreaves, K.R., Solonenko, N.E., Markillie, L.M., White, R.A., Brewer, H.M., Ansong, C., Orr, G., Adkins, J.N., and Sullivan, M.B. (2018). Multiple mechanisms drive phage infection efficiency in nearly identical hosts. ISME J. 12, 1605–1618.

Hoyles, L., McCartney, A.L., Neve, H., Gibson, G.R., Sanderson, J.D., Heller, K.J., and van Sinderen, D. (2014). Characterization of virus-like particles associated with the human faecal and caecal microbiota. Res. Microbiol. 165, 803–812.
Hryckowian, A.J., Merrill, B.D., Porter, N.T., Van Treuren, W., Nelson, E.J., Garlena, R.A., Russell, D.A., Martens, E.C., and Sonnenburg, J.L. (2020). Bacteroides thetaiotaomicron-Infecting Bacteriophage Isolates Inform Sequence-Based Host Range Predictions. Cell Host Microbe.

Huerta-Cepas, J., Forslund, K., Coelho, L.P., Szklarczyk, D., Jensen, L.J., von Mering, C., and Bork, P. (2017). Fast Genome-Wide Functional Annotation through Orthology Assignment by eggNOG-Mapper. Mol. Biol. Evol. 34, 2115–2122.

Hulo, C., de Castro, E., Masson, P., Bougueleret, L., Bairoch, A., Xenarios, I., and Le Mercier, P. (2011). ViralZone: a knowledge resource to understand virus diversity. Nucleic Acids Res. 39, D576-582.

Hyatt, D., Chen, G.-L., Locascio, P.F., Land, M.L., Larimer, F.W., and Hauser, L.J. (2010). Prodigal: prokaryotic gene recognition and translation initiation site identification. BMC Bioinformatics 11, 119.

Hyman, P., and Abedon, S.T. (2010). Bacteriophage host range and bacterial resistance. Adv. Appl. Microbiol. 70, 217–248.

Jahn, M.T., Arkhipova, K., Markert, S.M., Stigloher, C., Lachnit, T., Pita, L., Kupczok, A., Ribes, M., Stengel, S.T., Rosenstiel, P., et al. (2019). A Phage Protein Aids Bacterial Symbionts in Eukaryote Immune Evasion. Cell Host Microbe 26, 542-550.e5.

Jiang, X., Hall, A.B., Xavier, R.J., and Alm, E.J. (2019). Comprehensive analysis of chromosomal mobile genetic elements in the gut microbiome reveals phylum-level niche-adaptive gene pools. PLoS ONE 14, e0223680.

Karginov, F.V., and Hannon, G.J. (2010). The CRISPR system: small RNA-guided defense in bacteria and archaea. Mol. Cell 37, 7.

Katoh, K., Misawa, K., Kuma, K., and Miyata, T. (2002). MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Res. 30, 3059–3066.

Kelly, L., Ding, H., Huang, K.H., Osburne, M.S., and Chisholm, S.W. (2013). Genetic diversity in cultured and wild marine cyanomyoviruses reveals phosphorus stress as a strong selective agent. ISME J. 7, 1827–1841.

Keohane, D.M., Ghosh, T.S., Jeffery, I.B., Molloy, M.G., O'Toole, P.W., and Shanahan, F. (2020). Microbiome and health implications for ethnic minorities after enforced lifestyle changes. Nat. Med. 26, 1089–1095.

Kim, M.-S., and Bae, J.-W. (2018). Lysogeny is prevalent and widely distributed in the murine gut microbiota. ISME J. 12, 1127–1141.

Kim, M.-S., Park, E.-J., Roh, S.W., and Bae, J.-W. (2011). Diversity and Abundance of Single-Stranded DNA Viruses in Human Feces. Appl. Environ. Microbiol. 77, 8062–8070.

Kim, S., Ryu, K., Biswas, D., and Ahn, J. (2014). Survival, prophage induction, and invasive properties of lysogenic Salmonella Typhimurium exposed to simulated gastrointestinal conditions. Arch. Microbiol. 196, 655–659.

Knowles, B., Silveira, C.B., Bailey, B.A., Barott, K., Cantu, V.A., Cobián-Güemes, A.G., Coutinho, F.H., Dinsdale, E.A., Felts, B., Furby, K.A., et al. (2016). Lytic to temperate switching of viral communities. Nature 531, 466–470.

Koert, M., Mattson, C., Caruso, S., and Erill, I. (2019). Evidence for shared ancestry between Actinobacteria and Firmicutes bacteriophages. BioRxiv 842583.

Koonin, E.V., Dolja, V.V., Krupovic, M., Varsani, A., Wolf, Y.I., Yutin, N., Zerbini, F.M., and Kuhn, J.H. (2020). Global Organization and Proposed Megataxonomy of the Virus World. Microbiol. Mol. Biol. Rev. 84.

Krupovic, M., and Forterre, P. (2011). Microviridae Goes Temperate: Microvirus-Related Proviruses Reside in the Genomes of Bacteroidetes. PLOS ONE 6, e19893.

Lavigne, R., Darius, P., Summer, E.J., Seto, D., Mahadevan, P., Nilsson, A.S., Ackermann, H.W., and Kropinski, A.M. (2009). Classification of Myoviridae bacteriophages using protein sequence similarity. BMC Microbiol. 9, 224.

Lavigne, R., Seto, D., Mahadevan, P., Ackermann, H.-W., and Kropinski, A.M. (2008). Unifying classical and molecular taxonomic classification: analysis of the Podoviridae using BLASTP-based tools. Res. Microbiol. 159, 406–414.

Leinonen, R., Akhtar, R., Birney, E., Bower, L., Cerdeno-Tárraga, A., Cheng, Y., Cleland, I., Faruque, N., Goodgame, N., Gibson, R., et al. (2011). The European Nucleotide Archive. Nucleic Acids Res 39, D28–D31.

Letunic, I., and Bork, P. (2007). Interactive Tree Of Life (iTOL): an online tool for phylogenetic tree display and annotation. Bioinformatics 23, 127–128.

Li, D., Liu, C.-M., Luo, R., Sadakane, K., and Lam, T.-W. (2015). MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. Bioinformatics 31, 1674–1676.

Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows–Wheeler transform. Bioinformatics 25, 1754–1760.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., and 1000 Genome Project Data Processing Subgroup (2009). The Sequence Alignment/Map format and SAMtools. Bioinformatics 25, 2078–2079.

Li, W., and Godzik, A. (2006). Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. Bioinformatics 22, 1658–1659.

Lim, E.S., Zhou, Y., Zhao, G., Bauer, I.K., Droit, L., Ndao, I.M., Warner, B.B., Tarr, P.I., Wang, D., and Holtz, L.R. (2015). Early life dynamics of the human gut virome and bacterial microbiome in infants. Nat. Med. 21, 1228–1234.

Liu, M., Deora, R., Doulatov, S.R., Gingery, M., Eiserling, F.A., Preston, A., Maskell, D.J., Simons, R.W., Cotter, P.A., Parkhill, J., et al. (2002). Reverse transcriptase-mediated tropism switching in Bordetella bacteriophage. Science 295, 2091–2094.

Loh, B., Kuhn, A., and Leptihn, S. (2019). The fascinating biology behind phage display: filamentous phage assembly. Mol. Microbiol. 111, 1132–1138.

Lugli, G.A., Milani, C., Turroni, F., Tremblay, D., Ferrario, C., Mancabelli, L., Duranti, S., Ward, D.V., Ossiprandi, M.C., Moineau, S., et al. (2016). Prophages of the genus Bifidobacterium as modulating agents of the infant gut microbiota. Environ. Microbiol. 18, 2196–2213.

Lwoff, A. (1953). LYSOGENY1. Bacteriol. Rev. 17, 269–337.

Manrique, P., Bolduc, B., Walk, S.T., Oost, J. van der, Vos, W.M. de, and Young, M.J. (2016). Healthy human gut phageome. Proc. Natl. Acad. Sci. 113, 10400–10405.

Manrique, P., Dills, M., and Young, M.J. (2017). The Human Gut Phage Community and Its Implications for Health and Disease. Viruses 9.

Marbouty, M., Thierry, A., and Koszul, R. (2020). Phages - bacteria interactions network of the healthy human gut (Microbiology).

Meier-Kolthoff, J.P., and Göker, M. (2017). VICTOR: genome-based phylogeny and classification of prokaryotic viruses. Bioinformatics 33, 3396–3404.

Meyer, J.R., Dobias, D.T., Weitz, J.S., Barrick, J.E., Quick, R.T., and Lenski, R.E. (2012). Repeatability and Contingency in the Evolution of a Key Innovation in Phage Lambda. Science 335, 428–432.

Millman, A., Bernheim, A., Stokar-Avihail, A., Fedorenko, T., Voichek, M., Leavitt, A., and Sorek, R. (2020). Bacterial retrons function in anti-phage defense. BioRxiv 2020.06.21.156273.

Minot, S., Bryson, A., Chehoud, C., Wu, G.D., Lewis, J.D., and Bushman, F.D. (2013). Rapid evolution of the human gut virome. Proc. Natl. Acad. Sci. U. S. A. 110, 12450–12455.

Minot, S., Grunberg, S., Wu, G.D., Lewis, J.D., and Bushman, F.D. (2012). Hypervariable loci in the human gut virome. Proc. Natl. Acad. Sci. U. S. A. 109, 3962–3966.

Minot, S., Sinha, R., Chen, J., Li, H., Keilbaugh, S.A., Wu, G.D., Lewis, J.D., and Bushman, F.D. (2011). The human gut virome: Inter-individual variation and dynamic response to diet. Genome Res. 21, 1616–1625.

Moreno-Gallego, J.L., Chou, S.-P., Di Rienzi, S.C., Goodrich, J.K., Spector, T., Bell, J.T., Youngblut, N., Hewson, I., Reyes, A., and Ley, R.E. (2019). Virome diversity correlates with intestinal microbiome diversity in adult monozygotic twins. Cell Host Microbe 25, 261-272.e5.

Morgan, A.D., Bonsall, M.B., and Buckling, A. (2010). Impact of Bacterial Mutation Rate on Coevolutionary Dynamics Between Bacteria and Phages. Evolution 64, 2980–2987.

Morse, M.L., Lederberg, E.M., and Lederberg, J. (1956). Transduction in Escherichia Coli K-12. Genetics 41, 142–156.

Moussa, S.H., Lawler, J.L., and Young, R. (2014). Genetic Dissection of T4 Lysis. J. Bacteriol. 196, 2201–2209.

Muyzer, G., and Stams, A.J.M. (2008). The ecology and biotechnology of sulphate-reducing bacteria. Nat. Rev. Microbiol. 6, 441–454.

Naka, K., Qi, D., Yonesaki, T., and Otsuka, Y. (2017). RnlB Antitoxin of the Escherichia coli RnlA-RnlB Toxin–Antitoxin Module Requires RNase HI for Inhibition of RnlA Toxin Activity. Toxins 9.

Nayfach, S., Camargo, A.P., Eloe-Fadrosh, E., Roux, S., and Kyrpides, N. (2020). CheckV: assessing the quality of metagenome-assembled viral genomes. BioRxiv 2020.05.06.081778.

Norman, J.M., Handley, S.A., Baldridge, M.T., Droit, L., Liu, C.Y., Keller, B.C., Kambal, A.,

Monaco, C.L., Zhao, G., Fleshner, P., et al. (2015). Disease-specific Alterations in the Enteric Virome in Inflammatory Bowel Disease. Cell 160, 447–460.

Ofir, G., Melamed, S., Sberro, H., Mukamel, Z., Silverman, S., Yaakov, G., Doron, S., and Sorek, R. (2018). DISARM is a widespread bacterial defence system with broad anti-phage activities. Nat. Microbiol. 3, 90–98.

Ohshima, Y., Schumacher-Perdreau, F., Peters, G., and Pulverer, G. (1988). The role of capsule as a barrier to bacteriophage adsorption in an encapsulated Staphylococcus simulans strain. Med. Microbiol. Immunol. (Berl.) 177, 229–233.

Oliveira, P.H., Touchon, M., and Rocha, E.P.C. (2014). The interplay of restriction-modification systems with mobile genetic elements and their prokaryotic hosts. Nucleic Acids Res. 42, 10618–10631.
Ondov, B.D., Treangen, T.J., Melsted, P., Mallonee, A.B., Bergman, N.H., Koren, S., and Phillippy, A.M. (2016). Mash: fast genome and metagenome distance estimation using MinHash. Genome Biology 17, 132.

Otsuka, Y., and Yonesaki, T. (2012). Dmd of bacteriophage T4 functions as an antitoxin against Escherichia coli LsoA and RnlA toxins. Mol. Microbiol. 83, 669–681.

Paez-Espino, D., Roux, S., Chen, I.-M.A., Palaniappan, K., Ratner, A., Chu, K., Huntemann, M., Reddy, T.B.K., Pons, J.C., Llabrés, M., et al. (2019). IMG/VR v.2.0: an integrated data management and analysis system for cultivated and environmental viral genomes. Nucleic Acids Res. 47, D678–D686.

Paez-Espino, D., Sharon, I., Morovic, W., Stahl, B., Thomas, B.C., Barrangou, R., and Banfield, J.F. (2015). CRISPR immunity drives rapid phage genome evolution in Streptococcus thermophilus. MBio 6.

Porter, N.T., Hryckowian, A.J., Merrill, B.D., Fuentes, J.J., Gardner, J.O., Glowacki, R.W.P., Singh, S., Crawford, R.D., Snitkin, E.S., Sonnenburg, J.L., et al. (2020). Phase-variable capsular polysaccharides and lipoproteins modify bacteriophage susceptibility in Bacteroides thetaiotaomicron. Nat. Microbiol. 1–12.

Price, M.N., Dehal, P.S., and Arkin, A.P. (2010). FastTree 2 – Approximately Maximum-Likelihood Trees for Large Alignments. PLoS One 5.

Ren, J., Ahlgren, N.A., Lu, Y.Y., Fuhrman, J.A., and Sun, F. (2017). VirFinder: a novel k-mer based tool for identifying viral sequences from assembled metagenomic data. Microbiome 5, 69.

Reyes-Robles, T., Dillard, R.S., Cairns, L.S., Silva-Valenzuela, C.A., Housman, M., Ali, A., Wright, E.R., and Camilli, A. (2018). Vibrio cholerae Outer Membrane Vesicles Inhibit Bacteriophage Infection. J. Bacteriol. 200.

Reyes, A., Haynes, M., Hanson, N., Angly, F.E., Heath, A.C., Rohwer, F., and Gordon, J.I. (2010a). Viruses in the fecal microbiota of monozygotic twins and their mothers. Nature 466, 334–338.

Reyes, A., Semenkovich, N.P., Whiteson, K., Rohwer, F., and Gordon, J.I. (2012). Going viral: next-generation sequencing applied to phage populations in the human gut. Nat. Rev. Microbiol. 10, 607–617.

Ripp, S., and Miller, R.V. (1997). The role of pseudolysogeny in bacteriophage-host interactions in a natural freshwater environment. Microbiology, 143, 2065–2070.

Rohwer, F., and Edwards, R. (2002). The Phage Proteomic Tree: a genome-based taxonomy for phage. J. Bacteriol. 184, 4529–4535.

Roux, S., Enault, F., Hurwitz, B.L., and Sullivan, M.B. (2015). VirSorter: mining viral signal from microbial genomic data. PeerJ 3.

Roux, S., Krupovic, M., Poulet, A., Debroas, D., and Enault, F. (2012). Evolution and Diversity of the Microviridae Viral Family through a Collection of 81 New Complete Genomes Assembled from Virome Reads. PLoS ONE 7.

Samson, J.E., Magadán, A.H., Sabri, M., and Moineau, S. (2013). Revenge of the phages: defeating bacterial defences. Nat. Rev. Microbiol. 11, 675–687.

Schillinger, T., and Zingler, N. (2012). The low incidence of diversity-generating retroelements in sequenced genomes. Mob. Genet. Elem. 2, 287–291.

Schwarzer, D., Buettner, F.F.R., Browning, C., Nazarov, S., Rabsch, W., Bethe, A., Oberbeck, A., Bowman, V.D., Stummeyer, K., Mühlenhoff, M., et al. (2012). A multivalent adsorption apparatus explains the broad host range of phage phi92: a comprehensive genomic and structural analysis. J. Virol. 86, 10384–10398.

Seed, K.D. (2015). Battling Phages: How Bacteria Defend against Viral Attack. PLoS Pathog. 11.

Seemann, T. (2014). Prokka: rapid prokaryotic genome annotation. Bioinformatics 30, 2068–2069.

Seguritan, V., Alves, N., Arnoult, M., Raymond, A., Lorimer, D., Burgin, A.B., Salamon, P., and Segall, A.M. (2012). Artificial Neural Networks Trained to Detect Viral and Phage Structural Proteins. PLoS Comput. Biol. 8.

Shkoporov, A.N., and Hill, C. (2019). Bacteriophages of the Human Gut: The "Known Unknown" of the Microbiome. Cell Host Microbe 25, 195–209.

Shkoporov, A.N., Clooney, A.G., Sutton, T.D.S., Ryan, F.J., Daly, K.M., Nolan, J.A., McDonnell, S.A., Khokhlova, E.V., Draper, L.A., Forde, A., et al. (2019). The Human Gut Virome Is Highly Diverse, Stable, and Individual Specific. Cell Host Microbe 26, 527-541.e5.

Shkoporov, A.N., Khokhlova, E.V., Fitzgerald, C.B., Stockdale, S.R., Draper, L.A., Ross, R.P., and Hill, C. (2018b). ΦCrAss001 represents the most abundant bacteriophage family in the human gut and infects Bacteroides intestinalis. Nat. Commun. 9, 4781.

Shkoporov, A.N., Ryan, F.J., Draper, L.A., Forde, A., Stockdale, S.R., Daly, K.M., McDonnell, S.A., Nolan, J.A., Sutton, T.D.S., Dalmasso, M., et al. (2018a). Reproducible protocols for metagenomic analysis of human faecal phageomes. Microbiome 6, 68.

Silveira, C.B., and Rohwer, F.L. (2016). Piggyback-the-Winner in host-associated microbial communities. Npj Biofilms Microbiomes 2, 1–5.

Simmonds, P., Adams, M.J., Benkő, M., Breitbart, M., Brister, J.R., Carstens, E.B., Davison, A.J., Delwart, E., Gorbalenya, A.E., Harrach, B., et al. (2017). Consensus statement: Virus taxonomy in the age of metagenomics. Nat. Rev. Microbiol. 15, 161–168.

Söding, J., Biegert, A., and Lupas, A.N. (2005). The HHpred interactive server for protein homology detection and structure prediction. Nucleic Acids Res. 33, W244–W248.

Stern, A., Mick, E., Tirosh, I., Sagy, O., and Sorek, R. (2012). CRISPR targeting reveals a reservoir of common phages associated with the human gut microbiome. Genome Res. 22, 1985–1994.

Sumby, P., and Smith, M.C.M. (2002). Genetics of the phage growth limitation (Pgl) system of Streptomyces coelicolor A3(2). Mol. Microbiol. 44, 489–500.

Tailford, L.E., Crost, E.H., Kavanaugh, D., and Juge, N. (2015). Mucin glycan foraging in the human gut microbiome. Front. Genet. 6, 81.

Tetz, G., Brown, S.M., Hao, Y., and Tetz, V. (2018). Parkinson's disease and bacteriophages as its overlooked contributors. Sci. Rep. 8, 10812.

Thingstad, T.F. (2000). Elements of a theory for the mechanisms controlling abundance, diversity, and biogeochemical role of lytic bacterial viruses in aquatic systems. Limnol. Oceanogr. 45, 1320–1328.

Van Valen, L. (1973). A new evolutionary law. Evol Theory 1:1-30

Vangay, P., Johnson, A.J., Ward, T.L., Al-Ghalith, G.A., Shields-Cutler, R.R., Hillmann, B.M., Lucas, S.K., Beura, L.K., Thompson, E.A., Till, L.M., et al. (2018). U.S. immigration westernizes the human gut microbiome. Cell 175, 962-972.e10.

Waller, A.S., Yamada, T., Kristensen, D.M., Kultima, J.R., Sunagawa, S., Koonin, E.V., and Bork, P. (2014). Classification and quantification of bacteriophage taxa in human gut metagenomes. ISME J. 8, 1391–1402.

Warwick-Dugdale, J., Buchholz, H.H., Allen, M.J., and Temperton, B. (2019). Host-hijacking and planktonic piracy: how phages command the microbial high seas. Virol. J. 16, 15.

Wittebole, X., De Roock, S., and Opal, S.M. (2014). A historical overview of bacteriophage therapy as an alternative to antibiotics for the treatment of bacterial pathogens. Virulence 5, 226–235.

Wu, G.D., Chen, J., Hoffmann, C., Bittinger, K., Chen, Y.-Y., Keilbaugh, S.A., Bewtra, M., Knights, D., Walters, W.A., Knight, R., et al. (2011). Linking Long-Term Dietary Patterns with Gut Microbial Enterotypes. Science 334, 105–108.

Yehl, K., Lemire, S., Yang, A.C., Ando, H., Mimee, M., Torres, M.D.T., de la Fuente-Nunez, C., and Lu, T.K. (2019). Engineering Phage Host-Range and Suppressing Bacterial Resistance through Phage Tail Fiber Mutagenesis. Cell 179, 459-469.e9.

Young, R. (1992). Bacteriophage lysis: mechanism and regulation. Microbiol. Rev. 56, 430–481.

Yu, Z., An, B., Ramshaw, J.A.M., and Brodsky, B. (2014). Bacterial collagen-like proteins that form triple-helical structures. J. Struct. Biol. 186, 451–461.

Zhang, X., McDaniel, A.D., Wolf, L.E., Keusch, G.T., Waldor, M.K., and Acheson, D.W. (2000).

Quinolone antibiotics induce Shiga toxin-encoding bacteriophages, toxin production, and death in mice. J. Infect. Dis. 181, 664–670.

Zinder, N.D., and Lederberg, J. (1952). GENETIC EXCHANGE IN SALMONELLA1. J. Bacteriol. 64, 679–699.

Appendices

Appendix 1. Predicted hosts of the crAss-like family

The table include a list of bacterial species targeted by members of the crAss-like family.

crAss-like genus	Predicted hosts
I	NA
II	NA
III	*Bacteroides*_B *vulgatus*
IV	*Bacteroides xylanisolvens*_B
	Bacteroides caccae
V	*Fusicatenibacter saccharivorans*
	Lachnospira eligens
VI	*Bacteroides xylanisolvens*_B
	Prevotella copri
VII	*Bacteroides thetaiotaomicron*
	*Bacteroides*_B *massiliensis*
	*Bacteroides*_B *dorei*
	Bacteroides caccae
	Bacteroides faecis
	Bacteroides eggerthii
	*Bacteroides xylanisolvens*_B
	*Bacteroides*_B *vulgatus*
	Bacteroides uniformis
VIII	*Prevotella copri*
IX	*Prevotella copri*
X	*Parabacteroides merdae*
	Parabacteroides distasonis

Appendix 2. Metadata of deeply sequenced samples

The table include the metadata of the 3011 samples that were deeply sequenced (>50 million reads)

Run	Read count	Sample	Study	Health state	Lifestyle	Country	Continent
ERR209245	56435606	ERS199086	ERP002061	Diseased	Urban	Denmark	Europe
ERR209254	57112350	ERS199089	ERP002061	Diseased	Urban	Denmark	Europe
ERR209388	53374138	ERS199116	ERP002061	Diseased	Urban	Denmark	Europe
ERR209451	55720848	ERS199132	ERP002061	Diseased	Urban	Denmark	Europe
ERR209452	69094398	ERS199133	ERP002061	Healthy	Urban	Denmark	Europe
ERR209453	67499556	ERS199134	ERP002061	Diseased	Urban	Denmark	Europe
ERR209454	67416508	ERS199135	ERP002061	Diseased	Urban	Denmark	Europe
ERR209455	63468796	ERS199136	ERP002061	Healthy	Urban	Denmark	Europe
ERR209456	64862466	ERS199137	ERP002061	Healthy	Urban	Denmark	Europe
ERR209457	67290256	ERS199138	ERP002061	Diseased	Urban	Denmark	Europe
ERR209460	54265440	ERS199140	ERP002061	Healthy	Urban	Denmark	Europe
ERR209469	51902206	ERS199143	ERP002061	Healthy	Urban	Denmark	Europe
ERR209470	56314714	ERS199144	ERP002061	Healthy	Urban	Denmark	Europe
ERR209471	54016714	ERS199145	ERP002061	Diseased	Urban	Denmark	Europe
ERR209472	53545166	ERS199146	ERP002061	Diseased	Urban	Denmark	Europe
ERR209473	52538120	ERS199147	ERP002061	Diseased	Urban	Denmark	Europe
ERR209474	52347502	ERS199148	ERP002061	Diseased	Urban	Denmark	Europe
ERR209475	51747784	ERS199149	ERP002061	Diseased	Urban	Denmark	Europe
ERR209480	66091000	ERS199154	ERP002061	Diseased	Urban	Denmark	Europe
ERR209483	57676298	ERS199156	ERP002061	Healthy	Urban	Denmark	Europe
ERR209506	58554184	ERS199163	ERP002061	Diseased	Urban	Denmark	Europe
ERR209507	59509106	ERS199164	ERP002061	Healthy	NA	Denmark	Europe
ERR209508	57607172	ERS199165	ERP002061	Healthy	NA	Denmark	Europe
ERR209509	57292722	ERS199166	ERP002061	Healthy	NA	Denmark	Europe
ERR209514	57598544	ERS199168	ERP002061	Healthy	Urban	Denmark	Europe
ERR209515	59257266	ERS199169	ERP002061	Healthy	Urban	Denmark	Europe
ERR209516	52439256	ERS199170	ERP002061	Diseased	Urban	Denmark	Europe
ERR209517	65532334	ERS199171	ERP002061	Healthy	NA	Denmark	Europe
ERR209518	70607336	ERS199172	ERP002061	Diseased	Urban	Denmark	Europe
ERR209519	52406116	ERS199173	ERP002061	Healthy	NA	Denmark	Europe
ERR209527	70102972	ERS199175	ERP002061	Healthy	NA	Denmark	Europe
ERR209528	71269514	ERS199176	ERP002061	Diseased	Urban	Denmark	Europe
ERR209533	67249244	ERS199180	ERP002061	Diseased	Urban	Spain	Europe
ERR209536	61208458	ERS199178	ERP002061	Diseased	Urban	Spain	Europe
ERR209537	65467338	ERS199182	ERP002061	Diseased	Urban	Spain	Europe
ERR209540	60662794	ERS199186	ERP002061	Diseased	Urban	Spain	Europe
ERR209543	58204434	ERS199184	ERP002061	Diseased	Urban	Spain	Europe
ERR209546	63734438	ERS199188	ERP002061	Diseased	Urban	Spain	Europe
ERR209549	52539232	ERS199190	ERP002061	Diseased	Urban	Spain	Europe
ERR209553	50416090	ERS199192	ERP002061	Diseased	Urban	Spain	Europe
ERR209563	56337716	ERS199198	ERP002061	Diseased	Urban	Spain	Europe
ERR209566	52723686	ERS199200	ERP002061	Diseased	Urban	Spain	Europe
ERR209569	64869678	ERS199196	ERP002061	Diseased	Urban	Spain	Europe
ERR209574	50338214	ERS199204	ERP002061	Diseased	Urban	Spain	Europe
ERR209578	51782636	ERS199206	ERP002061	Diseased	Urban	Spain	Europe
ERR209579	62200606	ERS199207	ERP002061	Diseased	Urban	Spain	Europe
ERR209580	60482982	ERS199208	ERP002061	Diseased	Urban	Spain	Europe
ERR209581	57472814	ERS199209	ERP002061	Diseased	Urban	Spain	Europe
ERR209583	53396594	ERS199211	ERP002061	Diseased	Urban	Spain	Europe
ERR209587	50831192	ERS199215	ERP002061	Diseased	Urban	Spain	Europe
ERR209589	50124936	ERS199213	ERP002061	Diseased	Urban	Spain	Europe
ERR209599	55623438	ERS199219	ERP002061	Diseased	Urban	Spain	Europe
ERR209600	54325040	ERS199214	ERP002061	Diseased	Urban	Spain	Europe
ERR209603	58587784	ERS199221	ERP002061	Diseased	Urban	Spain	Europe
ERR209604	59533548	ERS199222	ERP002061	Diseased	Urban	Spain	Europe
ERR209606	58023176	ERS199224	ERP002061	Diseased	Urban	Spain	Europe
ERR209607	58532138	ERS199225	ERP002061	Diseased	Urban	Spain	Europe
ERR209608	57092478	ERS199226	ERP002061	Diseased	Urban	Spain	Europe
ERR209609	51588530	ERS199227	ERP002061	Diseased	Urban	Spain	Europe
ERR209611	54882022	ERS199228	ERP002061	Diseased	Urban	Spain	Europe
ERR209612	60271750	ERS199229	ERP002061	Diseased	Urban	Spain	Europe
ERR209613	53527386	ERS199232	ERP002061	Diseased	Urban	Spain	Europe
ERR209616	58716792	ERS199233	ERP002061	Diseased	Urban	Spain	Europe
ERR209617	56804664	ERS199234	ERP002061	Diseased	Urban	Spain	Europe
ERR209619	63963502	ERS199231	ERP002061	Diseased	Urban	Spain	Europe
ERR209620	52724784	ERS199236	ERP002061	Diseased	Urban	Spain	Europe
ERR209621	50698570	ERS199237	ERP002061	Diseased	Urban	Spain	Europe
ERR209623	50958600	ERS199238	ERP002061	Diseased	Urban	Spain	Europe
ERR209624	51536928	ERS199239	ERP002061	Diseased	Urban	Spain	Europe

ERR209625	56735036	ERS199240	ERP002061	Diseased	Urban	Spain	Europe
ERR209644	51598798	ERS199250	ERP002061	Healthy	Urban	Spain	Europe
ERR209648	51696896	ERS199251	ERP002061	Healthy	Urban	Spain	Europe
ERR209650	50957322	ERS199252	ERP002061	Healthy	Urban	Spain	Europe
ERR209651	58251944	ERS199253	ERP002061	Healthy	Urban	Spain	Europe
ERR209653	50853636	ERS199253	ERP002061	Healthy	Urban	Spain	Europe
ERR209654	58272584	ERS199254	ERP002061	Healthy	Urban	Spain	Europe
ERR209656	55227678	ERS199255	ERP002061	Healthy	Urban	Spain	Europe
ERR209659	57127536	ERS199257	ERP002061	Healthy	Urban	Spain	Europe
ERR209660	52792792	ERS199258	ERP002061	Healthy	Urban	Spain	Europe
ERR209661	52569744	ERS199259	ERP002061	Healthy	Urban	Spain	Europe
ERR209662	57805466	ERS199260	ERP002061	Healthy	Urban	Spain	Europe
ERR209663	61433124	ERS199261	ERP002061	Healthy	Urban	Spain	Europe
ERR209664	62337458	ERS199262	ERP002061	Healthy	Urban	Spain	Europe
ERR209665	56548044	ERS199263	ERP002061	Healthy	Urban	Spain	Europe
ERR209666	52998628	ERS199264	ERP002061	Healthy	Urban	Spain	Europe
ERR209667	55396106	ERS199265	ERP002061	Healthy	Urban	Spain	Europe
ERR209668	53091366	ERS199266	ERP002061	Healthy	Urban	Spain	Europe
ERR209669	54240216	ERS199267	ERP002061	Healthy	Urban	Spain	Europe
ERR209670	55202132	ERS199268	ERP002061	Healthy	Urban	Spain	Europe
ERR209671	55067876	ERS199269	ERP002061	Healthy	Urban	Spain	Europe
ERR209674	53085830	ERS199271	ERP002061	Healthy	Urban	Spain	Europe
ERR209678	52887150	ERS199273	ERP002061	Diseased	Urban	Spain	Europe
ERR209693	53234986	ERS199281	ERP002061	Diseased	Urban	Spain	Europe
ERR209694	54263644	ERS199282	ERP002061	Healthy	Urban	Spain	Europe
ERR209695	53810562	ERS199283	ERP002061	Diseased	Urban	Spain	Europe
ERR209706	50108648	ERS199289	ERP002061	Diseased	Urban	Spain	Europe
ERR209707	50591728	ERS199289	ERP002061	Diseased	Urban	Spain	Europe
ERR209714	55677194	ERS199293	ERP002061	Healthy	Urban	Spain	Europe
ERR209715	56192528	ERS199294	ERP002061	Diseased	Urban	Spain	Europe
ERR209716	56598978	ERS199295	ERP002061	Diseased	Urban	Spain	Europe
ERR209717	61335960	ERS199295	ERP002061	Diseased	Urban	Spain	Europe
ERR209718	54031568	ERS199296	ERP002061	Healthy	Urban	Spain	Europe
ERR209720	50334530	ERS199297	ERP002061	Healthy	Urban	Spain	Europe
ERR209721	58260832	ERS199298	ERP002061	Healthy	Urban	Spain	Europe
ERR209722	57920122	ERS199300	ERP002061	Healthy	Urban	Spain	Europe
ERR209736	52234832	ERS199306	ERP002061	Healthy	Urban	Spain	Europe
ERR209739	57615124	ERS199308	ERP002061	Diseased	Urban	Spain	Europe
ERR209740	59029856	ERS199308	ERP002061	Diseased	Urban	Spain	Europe
ERR209741	77193520	ERS199309	ERP002061	Healthy	Urban	Spain	Europe
ERR209742	76507678	ERS199309	ERP002061	Healthy	Urban	Spain	Europe
ERR209747	54762954	ERS199312	ERP002061	Diseased	Urban	Spain	Europe
ERR209756	57374224	ERS199316	ERP002061	Diseased	Urban	Spain	Europe
ERR209773	54010848	ERS199326	ERP002061	Diseased	Urban	Spain	Europe
ERR209802	54292856	ERS199341	ERP002061	Diseased	Urban	Spain	Europe
ERR209806	51730252	ERS199343	ERP002061	Diseased	Urban	Spain	Europe
ERR209809	52306110	ERS199345	ERP002061	Healthy	Urban	Spain	Europe
ERR209814	56417738	ERS199350	ERP002061	Healthy	Urban	Spain	Europe
ERR209815	54761572	ERS199348	ERP002061	Diseased	Urban	Spain	Europe
ERR209816	51844388	ERS199351	ERP002061	Diseased	Urban	Spain	Europe
ERR209832	51797806	ERS199358	ERP002061	Healthy	Urban	Spain	Europe
ERR209833	54439914	ERS199359	ERP002061	Healthy	Urban	Spain	Europe
ERR209837	51509446	ERS199361	ERP002061	Diseased	Urban	Spain	Europe
ERR209840	54550322	ERS199362	ERP002061	Diseased	Urban	Spain	Europe
ERR209843	55415504	ERS199366	ERP002061	Healthy	Urban	Spain	Europe
ERR209846	53935592	ERS199368	ERP002061	Healthy	Urban	Spain	Europe
ERR209852	53247724	ERS199370	ERP002061	Healthy	Urban	Spain	Europe
ERR209853	56209786	ERS199371	ERP002061	Diseased	Urban	Spain	Europe
ERR209875	50808912	ERS199376	ERP002061	Diseased	Urban	Spain	Europe
ERR209876	51002946	ERS199376	ERP002061	Diseased	Urban	Spain	Europe
ERR209898	51882940	ERS199391	ERP002061	Healthy	Urban	Spain	Europe
ERR321252	63915880	ERS328802	ERP003612	Diseased	Urban	Denmark	Europe
ERR321261	62462910	ERS328805	ERP003612	Healthy	Urban	Denmark	Europe
ERR321393	63363190	ERS328832	ERP003612	Diseased	Urban	Denmark	Europe
ERR321398	59394708	ERS328834	ERP003612	Diseased	Urban	Denmark	Europe
ERR321437	61160430	ERS328844	ERP003612	Healthy	Urban	Denmark	Europe
ERR321462	60425092	ERS328851	ERP003612	Diseased	Urban	Denmark	Europe
ERR321463	71831300	ERS328852	ERP003612	Healthy	Urban	Denmark	Europe
ERR321464	70427028	ERS328853	ERP003612	Healthy	Urban	Denmark	Europe
ERR321465	70759866	ERS328854	ERP003612	Healthy	Urban	Denmark	Europe
ERR321466	67012594	ERS328855	ERP003612	Healthy	Urban	Denmark	Europe
ERR321467	67997802	ERS328856	ERP003612	Healthy	Urban	Denmark	Europe

ERR321468	70338286	ERS328857	ERP003612	Healthy	Urban	Denmark	Europe
ERR321469	51249586	ERS328858	ERP003612	Healthy	Urban	Denmark	Europe
ERR321471	57592966	ERS328859	ERP003612	Healthy	Urban	Denmark	Europe
ERR321480	58869090	ERS328862	ERP003612	Healthy	Urban	Denmark	Europe
ERR321481	53940280	ERS328863	ERP003612	Healthy	Urban	Denmark	Europe
ERR321483	63513832	ERS328864	ERP003612	Healthy	Urban	Denmark	Europe
ERR321484	60898546	ERS328865	ERP003612	Healthy	Urban	Denmark	Europe
ERR321485	61157996	ERS328866	ERP003612	Healthy	Urban	Denmark	Europe
ERR321486	59646896	ERS328867	ERP003612	Healthy	Urban	Denmark	Europe
ERR321487	59466402	ERS328868	ERP003612	Healthy	Urban	Denmark	Europe
ERR321488	59688950	ERS328869	ERP003612	Diseased	Urban	Denmark	Europe
ERR321489	58429724	ERS328870	ERP003612	Healthy	Urban	Denmark	Europe
ERR321490	59409144	ERS328871	ERP003612	Healthy	Urban	Denmark	Europe
ERR321491	58836624	ERS328872	ERP003612	Healthy	Urban	Denmark	Europe
ERR321492	56395546	ERS328873	ERP003612	Healthy	Urban	Denmark	Europe
ERR321493	66564212	ERS328874	ERP003612	Healthy	Urban	Denmark	Europe
ERR321494	68547286	ERS328875	ERP003612	Healthy	Urban	Denmark	Europe
ERR321495	55003004	ERS328876	ERP003612	Healthy	Urban	Denmark	Europe
ERR321497	63093720	ERS328877	ERP003612	Healthy	Urban	Denmark	Europe
ERR321514	50516448	ERS328882	ERP003612	Healthy	Urban	Denmark	Europe
ERR321520	66181294	ERS328884	ERP003612	Healthy	Urban	Denmark	Europe
ERR321521	66887276	ERS328885	ERP003612	Healthy	Urban	Denmark	Europe
ERR321522	65626772	ERS328886	ERP003612	Healthy	Urban	Denmark	Europe
ERR321523	65331138	ERS328887	ERP003612	Healthy	Urban	Denmark	Europe
ERR321524	61901204	ERS328888	ERP003612	Healthy	Urban	Denmark	Europe
ERR321529	64217132	ERS328890	ERP003612	Healthy	Urban	Denmark	Europe
ERR321530	66357610	ERS328891	ERP003612	Healthy	Urban	Denmark	Europe
ERR321531	59658146	ERS328892	ERP003612	Healthy	Urban	Denmark	Europe
ERR321532	69823068	ERS328893	ERP003612	Healthy	Urban	Denmark	Europe
ERR321533	74372028	ERS328894	ERP003612	Healthy	Urban	Denmark	Europe
ERR321534	54332096	ERS328895	ERP003612	Healthy	Urban	Denmark	Europe
ERR321542	73808044	ERS328897	ERP003612	Healthy	Urban	Denmark	Europe
ERR321543	74388704	ERS328898	ERP003612	Diseased	Urban	Denmark	Europe
ERR321544	67283592	ERS328899	ERP003612	Healthy	Urban	Denmark	Europe
ERR321547	71018904	ERS328901	ERP003612	Healthy	Urban	Denmark	Europe
ERR321548	76050120	ERS328902	ERP003612	Healthy	Urban	Denmark	Europe
ERR321551	88854390	ERS328904	ERP003612	Healthy	Urban	Denmark	Europe
ERR321552	95619960	ERS328905	ERP003612	Healthy	Urban	Denmark	Europe
ERR321553	93076782	ERS328906	ERP003612	Healthy	Urban	Denmark	Europe
ERR321554	96817078	ERS328907	ERP003612	Healthy	Urban	Denmark	Europe
ERR321555	106852124	ERS328908	ERP003612	Healthy	Urban	Denmark	Europe
ERR321556	76670250	ERS328909	ERP003612	Healthy	Urban	Denmark	Europe
ERR321557	99289284	ERS328910	ERP003612	Healthy	Urban	Denmark	Europe
ERR321558	103827288	ERS328911	ERP003612	Healthy	Urban	Denmark	Europe
ERR321559	86385716	ERS328912	ERP003612	Healthy	Urban	Denmark	Europe
ERR321560	97854620	ERS328913	ERP003612	Healthy	Urban	Denmark	Europe
ERR321561	117066734	ERS328914	ERP003612	Healthy	Urban	Denmark	Europe
ERR321562	100321828	ERS328915	ERP003612	Healthy	Urban	Denmark	Europe
ERR321563	75873186	ERS328916	ERP003612	Healthy	Urban	Denmark	Europe
ERR321564	92747932	ERS328917	ERP003612	Healthy	Urban	Denmark	Europe
ERR321565	92940546	ERS328918	ERP003612	Healthy	Urban	Denmark	Europe
ERR321566	59598684	ERS328919	ERP003612	Healthy	Urban	Denmark	Europe
ERR321567	58274330	ERS328919	ERP003612	Healthy	Urban	Denmark	Europe
ERR321568	107338874	ERS328920	ERP003612	Healthy	Urban	Denmark	Europe
ERR321569	77164104	ERS328921	ERP003612	Healthy	Urban	Denmark	Europe
ERR321570	73553162	ERS328922	ERP003612	Healthy	Urban	Denmark	Europe
ERR321571	89250392	ERS328923	ERP003612	Healthy	Urban	Denmark	Europe
ERR321572	103705384	ERS328924	ERP003612	Healthy	Urban	Denmark	Europe
ERR321573	88973624	ERS328925	ERP003612	Healthy	Urban	Denmark	Europe
ERR321574	92188930	ERS328926	ERP003612	Healthy	Urban	Denmark	Europe
ERR321575	104778334	ERS328927	ERP003612	Healthy	Urban	Denmark	Europe
ERR321576	102848186	ERS328928	ERP003612	Healthy	Urban	Denmark	Europe
ERR321577	117735108	ERS328929	ERP003612	Healthy	Urban	Denmark	Europe
ERR321578	75367948	ERS328930	ERP003612	Healthy	Urban	Denmark	Europe
ERR321579	95332652	ERS328931	ERP003612	Healthy	Urban	Denmark	Europe
ERR321580	95313804	ERS328932	ERP003612	Healthy	Urban	Denmark	Europe
ERR321581	93178044	ERS328933	ERP003612	Healthy	Urban	Denmark	Europe
ERR321582	115683988	ERS328934	ERP003612	Healthy	Urban	Denmark	Europe
ERR321583	100315068	ERS328935	ERP003612	Healthy	Urban	Denmark	Europe
ERR321584	93420656	ERS328936	ERP003612	Healthy	Urban	Denmark	Europe
ERR321585	54369904	ERS328937	ERP003612	Healthy	Urban	Denmark	Europe
ERR321586	52998050	ERS328937	ERP003612	Healthy	Urban	Denmark	Europe

ERR321587	76108014	ERS328938	ERP003612	Healthy	Urban	Denmark	Europe
ERR321588	94435574	ERS328939	ERP003612	Healthy	Urban	Denmark	Europe
ERR321589	87272504	ERS328940	ERP003612	Healthy	Urban	Denmark	Europe
ERR321590	107825622	ERS328941	ERP003612	Healthy	Urban	Denmark	Europe
ERR321591	75586080	ERS328942	ERP003612	Healthy	Urban	Denmark	Europe
ERR321592	88723696	ERS328943	ERP003612	Healthy	Urban	Denmark	Europe
ERR321593	72051324	ERS328944	ERP003612	Healthy	Urban	Denmark	Europe
ERR321594	89611704	ERS328945	ERP003612	Healthy	Urban	Denmark	Europe
ERR321595	71564400	ERS328946	ERP003612	Healthy	Urban	Denmark	Europe
ERR321596	88148550	ERS328947	ERP003612	Healthy	Urban	Denmark	Europe
ERR321597	88681082	ERS328948	ERP003612	Healthy	Urban	Denmark	Europe
ERR321598	90135120	ERS328949	ERP003612	Healthy	Urban	Denmark	Europe
ERR321599	89840238	ERS328950	ERP003612	Healthy	Urban	Denmark	Europe
ERR321600	80140744	ERS328951	ERP003612	Healthy	Urban	Denmark	Europe
ERR321601	87625872	ERS328952	ERP003612	Healthy	Urban	Denmark	Europe
ERR321602	80453800	ERS328953	ERP003612	Healthy	Urban	Denmark	Europe
ERR321603	85952788	ERS328954	ERP003612	Healthy	Urban	Denmark	Europe
ERR321604	92775624	ERS328955	ERP003612	Healthy	Urban	Denmark	Europe
ERR321605	93457578	ERS328956	ERP003612	Healthy	Urban	Denmark	Europe
ERR321606	96057212	ERS328957	ERP003612	Healthy	Urban	Denmark	Europe
ERR321607	84908988	ERS328958	ERP003612	Healthy	Urban	Denmark	Europe
ERR321608	81239906	ERS328959	ERP003612	Healthy	Urban	Denmark	Europe
ERR321609	93478090	ERS328960	ERP003612	Healthy	Urban	Denmark	Europe
ERR321610	87068936	ERS328961	ERP003612	Healthy	Urban	Denmark	Europe
ERR321611	80848150	ERS328962	ERP003612	Healthy	Urban	Denmark	Europe
ERR321612	77472300	ERS328963	ERP003612	Healthy	Urban	Denmark	Europe
ERR321613	80011728	ERS328964	ERP003612	Healthy	Urban	Denmark	Europe
ERR321614	91695516	ERS328965	ERP003612	Healthy	Urban	Denmark	Europe
ERR321615	98884974	ERS328966	ERP003612	Healthy	Urban	Denmark	Europe
ERR321616	88543770	ERS328967	ERP003612	Healthy	Urban	Denmark	Europe
ERR321617	101560248	ERS328968	ERP003612	Healthy	Urban	Denmark	Europe
ERR321618	85828048	ERS328969	ERP003612	Healthy	Urban	Denmark	Europe
ERR321619	88583124	ERS328970	ERP003612	Healthy	Urban	Denmark	Europe
ERR321620	79482320	ERS328971	ERP003612	Healthy	Urban	Denmark	Europe
ERR321621	82395456	ERS328972	ERP003612	Healthy	Urban	Denmark	Europe
ERR321622	79036038	ERS328973	ERP003612	Healthy	Urban	Denmark	Europe
ERR321623	92673880	ERS328974	ERP003612	Healthy	Urban	Denmark	Europe
ERR321624	91294522	ERS328975	ERP003612	Healthy	Urban	Denmark	Europe
ERR321625	92421696	ERS328976	ERP003612	Healthy	Urban	Denmark	Europe
ERR321626	99861330	ERS328977	ERP003612	Healthy	Urban	Denmark	Europe
ERR321627	64860322	ERS328978	ERP003612	Healthy	Urban	Denmark	Europe
ERR321628	64825116	ERS328979	ERP003612	Healthy	Urban	Denmark	Europe
ERR321629	89069102	ERS328980	ERP003612	Healthy	Urban	Denmark	Europe
ERR321630	91767488	ERS328981	ERP003612	Healthy	Urban	Denmark	Europe
ERR321631	84668092	ERS328982	ERP003612	Healthy	Urban	Denmark	Europe
ERR321632	94429580	ERS328983	ERP003612	Healthy	Urban	Denmark	Europe
ERR321633	106319324	ERS328984	ERP003612	Healthy	Urban	Denmark	Europe
ERR321634	102315606	ERS328985	ERP003612	Healthy	Urban	Denmark	Europe
ERR321635	98666990	ERS328986	ERP003612	Healthy	Urban	Denmark	Europe
ERR321636	96413258	ERS328987	ERP003612	Healthy	Urban	Denmark	Europe
ERR321637	98064204	ERS328988	ERP003612	Healthy	Urban	Denmark	Europe
ERR321638	96794646	ERS328989	ERP003612	Healthy	Urban	Denmark	Europe
ERR321639	90292896	ERS328990	ERP003612	Healthy	Urban	Denmark	Europe
ERR321640	98897552	ERS328991	ERP003612	Healthy	Urban	Denmark	Europe
ERR321641	97466924	ERS328992	ERP003612	Healthy	Urban	Denmark	Europe
ERR321642	98712780	ERS328993	ERP003612	Diseased	Urban	Denmark	Europe
ERR321643	92037010	ERS328994	ERP003612	Healthy	Urban	Denmark	Europe
ERR321644	117074346	ERS328995	ERP003612	Healthy	Urban	Denmark	Europe
ERR321645	109244906	ERS328996	ERP003612	Healthy	Urban	Denmark	Europe
ERR321646	90147200	ERS328997	ERP003612	Healthy	Urban	Denmark	Europe
ERR321647	91964818	ERS328998	ERP003612	Healthy	Urban	Denmark	Europe
ERR321648	85941302	ERS328999	ERP003612	Healthy	Urban	Denmark	Europe
ERR321649	99801936	ERS329000	ERP003612	Healthy	Urban	Denmark	Europe
ERR321650	86737890	ERS329001	ERP003612	Healthy	Urban	Denmark	Europe
ERR321651	89107084	ERS329002	ERP003612	Healthy	Urban	Denmark	Europe
ERR321652	98098766	ERS329003	ERP003612	Healthy	Urban	Denmark	Europe
ERR321653	88114284	ERS329004	ERP003612	Healthy	Urban	Denmark	Europe
ERR321654	95130010	ERS329005	ERP003612	Healthy	Urban	Denmark	Europe
ERR321655	87603926	ERS329006	ERP003612	Healthy	Urban	Denmark	Europe
ERR321656	104533724	ERS329007	ERP003612	Healthy	Urban	Denmark	Europe
ERR414232	59286464	ERS396294	ERP004605	Healthy	Urban	Denmark	Europe
ERR414233	61410436	ERS396295	ERP004605	Healthy	Urban	Denmark	Europe

ERR414234	66407262	ERS396296	ERP004605	Healthy	Urban	Denmark	Europe
ERR414235	59380564	ERS396297	ERP004605	Healthy	Urban	Denmark	Europe
ERR414236	57735312	ERS396298	ERP004605	Healthy	Urban	Denmark	Europe
ERR414239	65588622	ERS396300	ERP004605	Healthy	Urban	Denmark	Europe
ERR414240	61378424	ERS396301	ERP004605	Healthy	Urban	Denmark	Europe
ERR414241	79286484	ERS396302	ERP004605	Healthy	Urban	Denmark	Europe
ERR414242	59402284	ERS396303	ERP004605	Healthy	Urban	Denmark	Europe
ERR414243	56766826	ERS396304	ERP004605	Healthy	Urban	Denmark	Europe
ERR414244	61201792	ERS396305	ERP004605	Healthy	Urban	Denmark	Europe
ERR414245	67189742	ERS396306	ERP004605	Healthy	Urban	Denmark	Europe
ERR414246	73561898	ERS396307	ERP004605	Healthy	Urban	Denmark	Europe
ERR414247	70006494	ERS396308	ERP004605	Healthy	Urban	Denmark	Europe
ERR414248	57118426	ERS396309	ERP004605	Healthy	Urban	Denmark	Europe
ERR414251	69579796	ERS396311	ERP004605	Healthy	Urban	Denmark	Europe
ERR414252	64696370	ERS396312	ERP004605	Healthy	Urban	Denmark	Europe
ERR414255	69218590	ERS396314	ERP004605	Healthy	Urban	Denmark	Europe
ERR414256	54771426	ERS396315	ERP004605	Healthy	Urban	Denmark	Europe
ERR414259	51043096	ERS396317	ERP004605	Healthy	Urban	Denmark	Europe
ERR414261	51047700	ERS396318	ERP004605	Healthy	Urban	Denmark	Europe
ERR414262	55485448	ERS396319	ERP004605	Healthy	Urban	Denmark	Europe
ERR414263	51951870	ERS396320	ERP004605	Healthy	Urban	Denmark	Europe
ERR414264	55779090	ERS396321	ERP004605	Healthy	Urban	Denmark	Europe
ERR414265	65958780	ERS396322	ERP004605	Healthy	Urban	Denmark	Europe
ERR414266	62520446	ERS396323	ERP004605	Healthy	Urban	Denmark	Europe
ERR414267	68258714	ERS396324	ERP004605	Healthy	Urban	Denmark	Europe
ERR414268	61753392	ERS396325	ERP004605	Healthy	Urban	Denmark	Europe
ERR414273	63983430	ERS396328	ERP004605	Healthy	Urban	Denmark	Europe
ERR414274	61770310	ERS396329	ERP004605	Healthy	Urban	Denmark	Europe
ERR414275	57088878	ERS396330	ERP004605	Healthy	Urban	Denmark	Europe
ERR414276	70247866	ERS396331	ERP004605	Healthy	Urban	Denmark	Europe
ERR414277	56120610	ERS396332	ERP004605	Healthy	Urban	Denmark	Europe
ERR414278	55010034	ERS396333	ERP004605	Healthy	Urban	Denmark	Europe
ERR414279	66286624	ERS396334	ERP004605	Healthy	Urban	Denmark	Europe
ERR414280	65432424	ERS396335	ERP004605	Healthy	Urban	Denmark	Europe
ERR414281	69637240	ERS396336	ERP004605	Healthy	Urban	Denmark	Europe
ERR414282	61744048	ERS396337	ERP004605	Healthy	Urban	Denmark	Europe
ERR414283	57576152	ERS396338	ERP004605	Healthy	Urban	Denmark	Europe
ERR414284	57033288	ERS396339	ERP004605	Healthy	Urban	Denmark	Europe
ERR414285	52485686	ERS396340	ERP004605	Healthy	Urban	Denmark	Europe
ERR414287	62229538	ERS396341	ERP004605	Healthy	Urban	Denmark	Europe
ERR414288	71766486	ERS396342	ERP004605	Healthy	Urban	Denmark	Europe
ERR414289	64207872	ERS396343	ERP004605	Healthy	Urban	Denmark	Europe
ERR414290	74846340	ERS396344	ERP004605	Healthy	Urban	Denmark	Europe
ERR414291	73252536	ERS396345	ERP004605	Healthy	Urban	Denmark	Europe
ERR414292	63447136	ERS396346	ERP004605	Healthy	Urban	Denmark	Europe
ERR414293	76256644	ERS396347	ERP004605	Healthy	Urban	Denmark	Europe
ERR414294	64616380	ERS396348	ERP004605	Healthy	Urban	Denmark	Europe
ERR414295	53244714	ERS396349	ERP004605	Healthy	Urban	Denmark	Europe
ERR414296	54005810	ERS396350	ERP004605	Healthy	Urban	Denmark	Europe
ERR414297	62401138	ERS396351	ERP004605	Healthy	Urban	Denmark	Europe
ERR414298	60977692	ERS396352	ERP004605	Healthy	Urban	Denmark	Europe
ERR414301	61628092	ERS396354	ERP004605	Healthy	Urban	Denmark	Europe
ERR414302	83141878	ERS396355	ERP004605	Healthy	Urban	Denmark	Europe
ERR414303	63367326	ERS396356	ERP004605	Healthy	Urban	Denmark	Europe
ERR414304	65939606	ERS396357	ERP004605	Healthy	Urban	Denmark	Europe
ERR414305	63052286	ERS396358	ERP004605	Healthy	Urban	Denmark	Europe
ERR414306	62532806	ERS396359	ERP004605	Healthy	Urban	Denmark	Europe
ERR414307	65773686	ERS396360	ERP004605	Healthy	Urban	Denmark	Europe
ERR414308	75579522	ERS396361	ERP004605	Healthy	Urban	Denmark	Europe
ERR414309	71450118	ERS396362	ERP004605	Healthy	Urban	Denmark	Europe
ERR414310	70280122	ERS396363	ERP004605	Healthy	Urban	Denmark	Europe
ERR414311	65259976	ERS396364	ERP004605	Healthy	Urban	Denmark	Europe
ERR414312	62522748	ERS396365	ERP004605	Healthy	Urban	Denmark	Europe
ERR414313	57558220	ERS396366	ERP004605	Healthy	Urban	Denmark	Europe
ERR414314	72060732	ERS396367	ERP004605	Healthy	Urban	Denmark	Europe
ERR414315	64835106	ERS396368	ERP004605	Healthy	Urban	Denmark	Europe
ERR414316	68199712	ERS396369	ERP004605	Healthy	Urban	Denmark	Europe
ERR414317	68408430	ERS396370	ERP004605	Healthy	Urban	Denmark	Europe
ERR414318	72479146	ERS396371	ERP004605	Healthy	Urban	Denmark	Europe
ERR414319	67284612	ERS396372	ERP004605	Healthy	Urban	Denmark	Europe
ERR414320	65878764	ERS396373	ERP004605	Healthy	Urban	Denmark	Europe
ERR414321	68748174	ERS396374	ERP004605	Healthy	Urban	Denmark	Europe

ERR414322	52542282	ERS396375	ERP004605	Healthy	Urban	Denmark	Europe
ERR414323	51522502	ERS396376	ERP004605	Healthy	Urban	Denmark	Europe
ERR414325	64787684	ERS396377	ERP004605	Healthy	Urban	Denmark	Europe
ERR414326	78865400	ERS396378	ERP004605	Healthy	Urban	Denmark	Europe
ERR414327	70110642	ERS396379	ERP004605	Healthy	Urban	Denmark	Europe
ERR414330	59395628	ERS396381	ERP004605	Healthy	Urban	Denmark	Europe
ERR414331	65252384	ERS396382	ERP004605	Healthy	Urban	Denmark	Europe
ERR414332	71882486	ERS396383	ERP004605	Healthy	Urban	Denmark	Europe
ERR414333	71937156	ERS396384	ERP004605	Healthy	Urban	Denmark	Europe
ERR414334	64551352	ERS396385	ERP004605	Healthy	Urban	Denmark	Europe
ERR414335	68663710	ERS396386	ERP004605	Healthy	Urban	Denmark	Europe
ERR414336	60671122	ERS396387	ERP004605	Healthy	Urban	Denmark	Europe
ERR414337	63798824	ERS396388	ERP004605	Healthy	Urban	Denmark	Europe
ERR414338	61863264	ERS396389	ERP004605	Healthy	Urban	Denmark	Europe
ERR414339	54655454	ERS396390	ERP004605	Healthy	Urban	Denmark	Europe
ERR414340	72458148	ERS396391	ERP004605	Healthy	Urban	Denmark	Europe
ERR414341	63285350	ERS396392	ERP004605	Healthy	Urban	Denmark	Europe
ERR414342	69995298	ERS396393	ERP004605	Healthy	Urban	Denmark	Europe
ERR414343	75252364	ERS396394	ERP004605	Healthy	Urban	Denmark	Europe
ERR414344	62254100	ERS396395	ERP004605	Healthy	Urban	Denmark	Europe
ERR414345	67023548	ERS396396	ERP004605	Healthy	Urban	Denmark	Europe
ERR414346	61655480	ERS396397	ERP004605	Healthy	Urban	Denmark	Europe
ERR414347	77866766	ERS396398	ERP004605	Healthy	Urban	Denmark	Europe
ERR414348	65665082	ERS396399	ERP004605	Healthy	Urban	Denmark	Europe
ERR414349	50392064	ERS396400	ERP004605	Healthy	Urban	Denmark	Europe
ERR414351	60351986	ERS396401	ERP004605	Healthy	Urban	Denmark	Europe
ERR414352	83312388	ERS396402	ERP004605	Healthy	Urban	Spain	Europe
ERR414353	75068128	ERS396403	ERP004605	Healthy	Urban	Spain	Europe
ERR414354	64245872	ERS396404	ERP004605	Healthy	Urban	Spain	Europe
ERR414355	89793772	ERS396405	ERP004605	Healthy	Urban	Spain	Europe
ERR414356	65287960	ERS396406	ERP004605	Healthy	Urban	Spain	Europe
ERR414357	78656946	ERS396408	ERP004605	Healthy	Urban	Spain	Europe
ERR414358	63612472	ERS396407	ERP004605	Healthy	Urban	Spain	Europe
ERR414359	68775918	ERS396409	ERP004605	Healthy	Urban	Spain	Europe
ERR414360	89772802	ERS396410	ERP004605	Healthy	Urban	Spain	Europe
ERR414361	63841736	ERS396411	ERP004605	Healthy	Urban	Spain	Europe
ERR414362	70376396	ERS396412	ERP004605	Healthy	Urban	Spain	Europe
ERR414363	52496332	ERS396413	ERP004605	Healthy	Urban	Spain	Europe
ERR414364	67667738	ERS396414	ERP004605	Healthy	Urban	Spain	Europe
ERR414365	79648216	ERS396415	ERP004605	Healthy	Urban	Spain	Europe
ERR414366	65644110	ERS396416	ERP004605	Healthy	Urban	Spain	Europe
ERR414367	91200006	ERS396417	ERP004605	Healthy	Urban	Spain	Europe
ERR414368	57089160	ERS396418	ERP004605	Healthy	Urban	Spain	Europe
ERR414369	64229756	ERS396419	ERP004605	Healthy	Urban	Spain	Europe
ERR414370	68513678	ERS396421	ERP004605	Healthy	Urban	Spain	Europe
ERR414371	67707784	ERS396420	ERP004605	Healthy	Urban	Spain	Europe
ERR414372	56828128	ERS396422	ERP004605	Healthy	Urban	Spain	Europe
ERR414374	70857344	ERS396424	ERP004605	Healthy	Urban	Spain	Europe
ERR414375	73321522	ERS396425	ERP004605	Healthy	Urban	Spain	Europe
ERR414376	57593656	ERS396426	ERP004605	Healthy	Urban	Spain	Europe
ERR414377	75092598	ERS396427	ERP004605	Healthy	Urban	Spain	Europe
ERR414378	59797984	ERS396428	ERP004605	Healthy	Urban	Spain	Europe
ERR414379	67592192	ERS396429	ERP004605	Healthy	Urban	Spain	Europe
ERR414380	60766656	ERS396430	ERP004605	Healthy	Urban	Spain	Europe
ERR414381	76459116	ERS396431	ERP004605	Healthy	Urban	Spain	Europe
ERR414382	69320298	ERS396433	ERP004605	Healthy	Urban	Spain	Europe
ERR414383	70285588	ERS396432	ERP004605	Healthy	Urban	Spain	Europe
ERR414384	68849288	ERS396434	ERP004605	Healthy	Urban	Spain	Europe
ERR414385	71137080	ERS396435	ERP004605	Healthy	Urban	Spain	Europe
ERR414388	63236984	ERS396437	ERP004605	Healthy	Urban	Spain	Europe
ERR414389	92869162	ERS396438	ERP004605	Healthy	Urban	Spain	Europe
ERR414390	79693966	ERS396439	ERP004605	Healthy	Urban	Spain	Europe
ERR414391	82421604	ERS396440	ERP004605	Healthy	Urban	Spain	Europe
ERR414392	84923280	ERS396442	ERP004605	Healthy	Urban	Spain	Europe
ERR414393	82678076	ERS396441	ERP004605	Healthy	Urban	Spain	Europe
ERR414394	76311838	ERS396443	ERP004605	Healthy	Urban	Spain	Europe
ERR414395	76484192	ERS396444	ERP004605	Healthy	Urban	Spain	Europe
ERR414396	89330254	ERS396445	ERP004605	Healthy	Urban	Spain	Europe
ERR414397	69023156	ERS396446	ERP004605	Healthy	Urban	Spain	Europe
ERR414398	103333716	ERS396447	ERP004605	Healthy	Urban	Spain	Europe
ERR414399	73243468	ERS396448	ERP004605	Healthy	Urban	Spain	Europe
ERR414400	74468102	ERS396449	ERP004605	Healthy	Urban	Spain	Europe

ERR414401	75064734	ERS396450	ERP004605	Healthy	Urban	Spain	Europe
ERR414402	61782086	ERS396451	ERP004605	Healthy	Urban	Spain	Europe
ERR414403	66073128	ERS396452	ERP004605	Healthy	Urban	Spain	Europe
ERR414404	58648450	ERS396453	ERP004605	Healthy	Urban	Spain	Europe
ERR414405	60799638	ERS396454	ERP004605	Healthy	Urban	Spain	Europe
ERR414406	61181112	ERS396455	ERP004605	Healthy	Urban	Spain	Europe
ERR414407	62688288	ERS396456	ERP004605	Healthy	Urban	Spain	Europe
ERR414408	74809564	ERS396457	ERP004605	Healthy	Urban	Spain	Europe
ERR414409	73492052	ERS396458	ERP004605	Healthy	Urban	Spain	Europe
ERR414410	74227630	ERS396459	ERP004605	Healthy	Urban	Spain	Europe
ERR414412	54448198	ERS396460	ERP004605	Healthy	Urban	Spain	Europe
ERR414413	72785188	ERS396461	ERP004605	Healthy	Urban	Spain	Europe
ERR414416	59333776	ERS396463	ERP004605	Healthy	Urban	Spain	Europe
ERR414417	90947146	ERS396464	ERP004605	Healthy	Urban	Spain	Europe
ERR414418	74543272	ERS396465	ERP004605	Healthy	Urban	Spain	Europe
ERR414419	114209870	ERS396466	ERP004605	Healthy	Urban	Spain	Europe
ERR414420	50638890	ERS396467	ERP004605	Healthy	Urban	Spain	Europe
ERR414424	70979726	ERS396469	ERP004605	Healthy	Urban	Spain	Europe
ERR414425	68094754	ERS396470	ERP004605	Healthy	Urban	Spain	Europe
ERR414428	64403446	ERS396472	ERP004605	Healthy	Urban	Spain	Europe
ERR414429	101382234	ERS396473	ERP004605	Healthy	Urban	Spain	Europe
ERR414430	64592722	ERS396474	ERP004605	Healthy	Urban	Spain	Europe
ERR414433	89157464	ERS396476	ERP004605	Healthy	Urban	Spain	Europe
ERR414440	69680040	ERS396480	ERP004605	Healthy	Urban	Spain	Europe
ERR414441	74419132	ERS396481	ERP004605	Healthy	Urban	Spain	Europe
ERR414442	68053564	ERS396482	ERP004605	Healthy	Urban	Spain	Europe
ERR414443	78457044	ERS396483	ERP004605	Healthy	Urban	Spain	Europe
ERR414444	83199388	ERS396484	ERP004605	Healthy	Urban	Spain	Europe
ERR414445	69344702	ERS396485	ERP004605	Healthy	Urban	Spain	Europe
ERR414446	84735326	ERS396486	ERP004605	Healthy	Urban	Spain	Europe
ERR414447	79247452	ERS396487	ERP004605	Healthy	Urban	Spain	Europe
ERR414448	67023734	ERS396488	ERP004605	Healthy	Urban	Spain	Europe
ERR414449	99463574	ERS396489	ERP004605	Healthy	Urban	Spain	Europe
ERR414450	68004794	ERS396490	ERP004605	Healthy	Urban	Spain	Europe
ERR414451	79490748	ERS396491	ERP004605	Healthy	Urban	Spain	Europe
ERR414452	80559990	ERS396492	ERP004605	Healthy	Urban	Spain	Europe
ERR414453	106262384	ERS396493	ERP004605	Healthy	Urban	Spain	Europe
ERR414454	79013374	ERS396494	ERP004605	Healthy	Urban	Spain	Europe
ERR414455	83651770	ERS396495	ERP004605	Healthy	Urban	Spain	Europe
ERR414456	70051112	ERS396496	ERP004605	Healthy	Urban	Spain	Europe
ERR414457	79808560	ERS396497	ERP004605	Healthy	Urban	Spain	Europe
ERR414458	71316094	ERS396498	ERP004605	Healthy	Urban	Spain	Europe
ERR414459	82993752	ERS396499	ERP004605	Healthy	Urban	Spain	Europe
ERR414460	76836956	ERS396500	ERP004605	Healthy	Urban	Spain	Europe
ERR414461	85054812	ERS396501	ERP004605	Healthy	Urban	Spain	Europe
ERR414462	73282812	ERS396502	ERP004605	Healthy	Urban	Spain	Europe
ERR414463	86563786	ERS396503	ERP004605	Healthy	Urban	Spain	Europe
ERR414464	76612664	ERS396504	ERP004605	Healthy	Urban	Spain	Europe
ERR414465	83905566	ERS396505	ERP004605	Healthy	Urban	Spain	Europe
ERR414466	81026748	ERS396506	ERP004605	Healthy	Urban	Spain	Europe
ERR414467	78344578	ERS396507	ERP004605	Healthy	Urban	Spain	Europe
ERR414468	72153214	ERS396508	ERP004605	Healthy	Urban	Spain	Europe
ERR414469	81937928	ERS396509	ERP004605	Healthy	Urban	Spain	Europe
ERR414470	75782416	ERS396510	ERP004605	Healthy	Urban	Spain	Europe
ERR414471	82023626	ERS396511	ERP004605	Healthy	Urban	Spain	Europe
ERR414472	56448616	ERS396512	ERP004605	Healthy	Urban	Spain	Europe
ERR414473	69524978	ERS396513	ERP004605	Healthy	Urban	Spain	Europe
ERR414474	61495840	ERS396514	ERP004605	Healthy	Urban	Spain	Europe
ERR414475	132964664	ERS396515	ERP004605	Healthy	Urban	Spain	Europe
ERR414476	75035580	ERS396516	ERP004605	Healthy	Urban	Spain	Europe
ERR414477	81529128	ERS396517	ERP004605	Healthy	Urban	Spain	Europe
ERR414478	66420338	ERS396518	ERP004605	Healthy	Urban	Spain	Europe
ERR414479	77121586	ERS396519	ERP004605	Healthy	Urban	Spain	Europe
ERR414480	57738776	ERS396520	ERP004605	Healthy	Urban	Spain	Europe
ERR414481	72102506	ERS396521	ERP004605	Healthy	Urban	Spain	Europe
ERR414482	61651536	ERS396522	ERP004605	Healthy	Urban	Spain	Europe
ERR414483	86307728	ERS396523	ERP004605	Healthy	Urban	Spain	Europe
ERR414484	78523990	ERS396524	ERP004605	Healthy	Urban	Spain	Europe
ERR414485	86384876	ERS396525	ERP004605	Healthy	Urban	Spain	Europe
ERR414486	64525152	ERS396526	ERP004605	Healthy	Urban	Spain	Europe
ERR414487	85391554	ERS396527	ERP004605	Healthy	Urban	Spain	Europe
ERR414488	80673146	ERS396528	ERP004605	Healthy	Urban	Spain	Europe

ERR414493	50686226	ERS396531	ERP004605	Healthy	Urban	Spain	Europe
ERR414496	63981644	ERS396533	ERP004605	Healthy	Urban	Spain	Europe
ERR414497	69707948	ERS396534	ERP004605	Healthy	Urban	Spain	Europe
ERR414500	70147014	ERS396536	ERP004605	Healthy	Urban	Spain	Europe
ERR414510	62751602	ERS396541	ERP004605	Healthy	Urban	Spain	Europe
ERR479083	61990744	ERS436744	ERP005534	Diseased	Urban	France	Europe
ERR479084	61873562	ERS436744	ERP005534	Diseased	Urban	France	Europe
ERR479254	117226114	ERS436712	ERP005534	Diseased	Urban	France	Europe
ERR479298	125852494	ERS436646	ERP005534	Diseased	Urban	France	Europe
ERR479302	71232782	ERS436703	ERP005534	Diseased	Urban	France	Europe
ERR479303	73853888	ERS436703	ERP005534	Diseased	Urban	France	Europe
ERR505083	58369688	ERS433542	ERP005558	Healthy	Urban	Australia	Oceania
ERR505084	75792852	ERS433543	ERP005558	Healthy	Urban	Australia	Oceania
ERR505085	81335886	ERS433544	ERP005558	Healthy	Urban	Australia	Oceania
ERR505087	61928528	ERS433546	ERP005558	Healthy	Urban	Australia	Oceania
ERR505088	61257354	ERS433547	ERP005558	Healthy	Urban	Australia	Oceania
ERR505089	61111978	ERS433548	ERP005558	Healthy	Urban	Australia	Oceania
ERR505090	61748134	ERS433549	ERP005558	Healthy	Urban	Australia	Oceania
ERR505091	69947174	ERS433550	ERP005558	Healthy	Urban	Australia	Oceania
ERR505092	85249004	ERS433551	ERP005558	Healthy	Urban	Australia	Oceania
ERR505093	76762980	ERS433552	ERP005558	Healthy	Urban	Australia	Oceania
ERR505094	74532720	ERS433553	ERP005558	Healthy	Urban	Australia	Oceania
ERR505098	52765762	ERS433557	ERP005558	Healthy	Urban	Australia	Oceania
ERR505102	55812886	ERS433561	ERP005558	Healthy	Urban	Australia	Oceania
ERR505103	62379908	ERS433562	ERP005558	Healthy	Urban	Australia	Oceania
ERR505104	55950870	ERS433563	ERP005558	Healthy	Urban	Australia	Oceania
ERR505105	54048916	ERS433564	ERP005558	Healthy	Urban	Australia	Oceania
ERR525700	56112014	ERS473035	ERP005989	Healthy	Urban	Denmark	Europe
ERR525709	59468788	ERS473044	ERP005989	Healthy	Urban	Denmark	Europe
ERR525712	54728592	ERS473047	ERP005989	Healthy	Urban	Denmark	Europe
ERR525772	51251734	ERS473107	ERP005989	Healthy	Urban	Denmark	Europe
ERR525804	50894984	ERS473139	ERP005989	Healthy	Urban	Denmark	Europe
ERR525810	53817202	ERS473145	ERP005989	Healthy	Urban	Denmark	Europe
ERR525813	57291452	ERS473148	ERP005989	Healthy	Urban	Denmark	Europe
ERR525817	56759370	ERS473152	ERP005989	Healthy	Urban	Denmark	Europe
ERR525841	78778224	ERS473176	ERP005989	NA	Urban	Denmark	Europe
ERR525856	59661654	ERS473191	ERP005989	Healthy	Urban	Denmark	Europe
ERR525863	51726294	ERS473198	ERP005989	Healthy	Urban	Denmark	Europe
ERR525868	54310576	ERS473203	ERP005989	Healthy	Urban	Denmark	Europe
ERR525871	55682174	ERS473206	ERP005989	Healthy	Urban	Denmark	Europe
ERR525876	51054204	ERS473211	ERP005989	Healthy	Urban	Denmark	Europe
ERR525896	50271380	ERS473231	ERP005989	Healthy	Urban	Denmark	Europe
ERR525898	51248116	ERS473233	ERP005989	Healthy	Urban	Denmark	Europe
ERR525907	51632000	ERS473242	ERP005989	Healthy	Urban	Denmark	Europe
ERR525912	51149768	ERS473247	ERP005989	Healthy	Urban	Denmark	Europe
ERR525915	59475128	ERS473250	ERP005989	Healthy	Urban	Denmark	Europe
ERR525923	54797298	ERS473258	ERP005989	Healthy	Urban	Denmark	Europe
ERR525924	52036610	ERS473259	ERP005989	Healthy	Urban	Denmark	Europe
ERR525926	50987964	ERS473261	ERP005989	Healthy	Urban	Denmark	Europe
ERR525927	50971764	ERS473262	ERP005989	Healthy	Urban	Denmark	Europe
ERR525928	60059318	ERS473263	ERP005989	Healthy	Urban	Denmark	Europe
ERR525935	55494752	ERS473270	ERP005989	Healthy	Urban	Denmark	Europe
ERR525936	56408186	ERS473271	ERP005989	Healthy	Urban	Denmark	Europe
ERR525938	59022986	ERS473273	ERP005989	Healthy	Urban	Denmark	Europe
ERR525946	55713120	ERS473281	ERP005989	Healthy	Urban	Denmark	Europe
ERR525970	53378800	ERS473305	ERP005989	Healthy	Urban	Denmark	Europe
ERR526004	50018352	ERS473339	ERP005989	Healthy	Urban	Denmark	Europe
ERR526016	53835110	ERS473351	ERP005989	Healthy	Urban	Denmark	Europe
ERR526056	56566616	ERS473391	ERP005989	NA	Urban	Denmark	Europe
ERR526059	50971338	ERS473394	ERP005989	NA	Urban	Denmark	Europe
ERR526064	54561744	ERS473399	ERP005989	Healthy	Urban	Denmark	Europe
ERR526066	56218302	ERS473401	ERP005989	Healthy	Urban	Denmark	Europe
ERR526067	51052244	ERS473402	ERP005989	Healthy	Urban	Denmark	Europe
ERR526070	50557510	ERS473405	ERP005989	Healthy	Urban	Denmark	Europe
ERR526071	52050994	ERS473406	ERP005989	Healthy	Urban	Denmark	Europe
ERR526081	57768688	ERS473416	ERP005989	Healthy	Urban	Denmark	Europe
ERR526084	51286372	ERS473419	ERP005989	Healthy	Urban	Denmark	Europe
ERR526086	55277092	ERS473421	ERP005989	Healthy	Urban	Denmark	Europe
ERR589440	57698296	ERS537179	ERP006678	NA	NA	China	Asia
ERR589441	58690164	ERS537180	ERP006678	NA	NA	China	Asia
ERR589442	58882664	ERS537181	ERP006678	NA	NA	China	Asia
ERR589443	57079552	ERS537182	ERP006678	NA	NA	China	Asia

ERR589448	51017916	ERS537183	ERP006678	NA	NA	China	Asia
ERR589452	59983100	ERS537184	ERP006678	NA	NA	China	Asia
ERR589485	54566326	ERS537185	ERP006678	NA	NA	China	Asia
ERR589501	54353102	ERS537186	ERP006678	NA	NA	China	Asia
ERR589502	55292064	ERS537187	ERP006678	NA	NA	China	Asia
ERR589503	55098830	ERS537188	ERP006678	NA	NA	China	Asia
ERR589504	53577076	ERS537189	ERP006678	NA	NA	China	Asia
ERR589505	74595622	ERS537190	ERP006678	NA	NA	China	Asia
ERR589506	64707260	ERS537191	ERP006678	NA	NA	China	Asia
ERR589507	50560142	ERS537192	ERP006678	NA	NA	China	Asia
ERR589512	65248448	ERS537197	ERP006678	NA	NA	China	Asia
ERR589513	58972426	ERS537198	ERP006678	NA	NA	China	Asia
ERR589527	76635200	ERS537199	ERP006678	NA	NA	China	Asia
ERR589528	56108844	ERS537200	ERP006678	NA	NA	China	Asia
ERR589529	50675110	ERS537201	ERP006678	NA	NA	China	Asia
ERR589530	69853664	ERS537202	ERP006678	NA	NA	China	Asia
ERR589531	66957738	ERS537203	ERP006678	NA	NA	China	Asia
ERR589532	50349028	ERS537204	ERP006678	NA	NA	China	Asia
ERR589533	57569974	ERS537205	ERP006678	NA	NA	China	Asia
ERR589534	51950850	ERS537206	ERP006678	NA	NA	China	Asia
ERR589535	60617238	ERS537207	ERP006678	NA	NA	China	Asia
ERR589537	63508230	ERS537209	ERP006678	NA	NA	China	Asia
ERR589538	68452656	ERS537210	ERP006678	NA	NA	China	Asia
ERR589539	60064360	ERS537211	ERP006678	NA	NA	China	Asia
ERR589540	60889366	ERS537212	ERP006678	NA	NA	China	Asia
ERR589542	74037834	ERS537213	ERP006678	NA	NA	China	Asia
ERR589543	61220320	ERS537214	ERP006678	NA	NA	China	Asia
ERR589544	55600256	ERS537215	ERP006678	NA	NA	China	Asia
ERR589545	59204452	ERS537216	ERP006678	NA	NA	China	Asia
ERR589546	78325926	ERS537217	ERP006678	NA	NA	China	Asia
ERR589548	74419642	ERS537219	ERP006678	NA	NA	China	Asia
ERR589549	54857414	ERS537220	ERP006678	NA	NA	China	Asia
ERR589550	51270852	ERS537221	ERP006678	NA	NA	China	Asia
ERR589552	59613916	ERS537223	ERP006678	NA	NA	China	Asia
ERR589553	62835924	ERS537224	ERP006678	NA	NA	China	Asia
ERR589554	63857998	ERS537225	ERP006678	NA	NA	China	Asia
ERR589555	66151710	ERS537226	ERP006678	NA	NA	China	Asia
ERR589556	73008518	ERS537227	ERP006678	NA	NA	China	Asia
ERR589557	66863232	ERS537228	ERP006678	NA	NA	China	Asia
ERR589558	65062888	ERS537229	ERP006678	NA	NA	China	Asia
ERR589559	57906998	ERS537230	ERP006678	NA	NA	China	Asia
ERR589560	65891066	ERS537231	ERP006678	NA	NA	China	Asia
ERR589561	69486914	ERS537232	ERP006678	NA	NA	China	Asia
ERR589562	59193110	ERS537233	ERP006678	NA	NA	China	Asia
ERR589563	57434138	ERS537234	ERP006678	NA	NA	China	Asia
ERR589564	73297002	ERS537235	ERP006678	NA	NA	China	Asia
ERR589565	63082320	ERS537236	ERP006678	NA	NA	China	Asia
ERR589566	61593814	ERS537237	ERP006678	NA	NA	China	Asia
ERR589567	63379336	ERS537238	ERP006678	NA	NA	China	Asia
ERR589568	58068006	ERS537239	ERP006678	NA	NA	China	Asia
ERR589569	63924258	ERS537240	ERP006678	NA	NA	China	Asia
ERR589570	57533064	ERS537241	ERP006678	NA	NA	China	Asia
ERR589571	63503624	ERS537242	ERP006678	NA	NA	China	Asia
ERR589572	72902988	ERS537243	ERP006678	NA	NA	China	Asia
ERR589573	75073360	ERS537244	ERP006678	NA	NA	China	Asia
ERR589574	67482080	ERS537245	ERP006678	NA	NA	China	Asia
ERR589575	65653170	ERS537246	ERP006678	NA	NA	China	Asia
ERR589576	57052002	ERS537247	ERP006678	NA	NA	China	Asia
ERR589577	61099476	ERS537248	ERP006678	NA	NA	China	Asia
ERR589578	59708870	ERS537249	ERP006678	NA	NA	China	Asia
ERR589579	62378556	ERS537250	ERP006678	NA	NA	China	Asia
ERR589580	67824848	ERS537251	ERP006678	NA	NA	China	Asia
ERR589581	69995556	ERS537252	ERP006678	NA	NA	China	Asia
ERR589582	70191886	ERS537253	ERP006678	NA	NA	China	Asia
ERR589583	68236492	ERS537254	ERP006678	NA	NA	China	Asia
ERR589584	59430170	ERS537255	ERP006678	NA	NA	China	Asia
ERR589585	67656686	ERS537256	ERP006678	NA	NA	China	Asia
ERR589586	63051324	ERS537257	ERP006678	NA	NA	China	Asia
ERR589587	66003504	ERS537258	ERP006678	NA	NA	China	Asia
ERR589588	71250322	ERS537259	ERP006678	NA	NA	China	Asia
ERR589589	58561834	ERS537260	ERP006678	NA	NA	China	Asia
ERR589590	71746422	ERS537261	ERP006678	NA	NA	China	Asia

ERR589591	78670720	ERS537262	ERP006678	NA	NA	China	Asia
ERR589592	66740048	ERS537263	ERP006678	NA	NA	China	Asia
ERR589593	70665438	ERS537264	ERP006678	NA	NA	China	Asia
ERR589594	62370300	ERS537265	ERP006678	NA	NA	China	Asia
ERR589595	63197186	ERS537266	ERP006678	NA	NA	China	Asia
ERR589732	71016022	ERS537267	ERP006678	NA	NA	China	Asia
ERR589733	76044138	ERS537268	ERP006678	NA	NA	China	Asia
ERR589734	78280332	ERS537269	ERP006678	NA	NA	China	Asia
ERR589735	79913064	ERS537270	ERP006678	NA	NA	China	Asia
ERR589736	69933476	ERS537271	ERP006678	NA	NA	China	Asia
ERR589737	76453834	ERS537272	ERP006678	NA	NA	China	Asia
ERR589738	63545170	ERS537273	ERP006678	NA	NA	China	Asia
ERR589739	73144578	ERS537274	ERP006678	NA	NA	China	Asia
ERR589740	73389146	ERS537275	ERP006678	NA	NA	China	Asia
ERR589741	77623188	ERS537276	ERP006678	NA	NA	China	Asia
ERR589742	62465476	ERS537277	ERP006678	NA	NA	China	Asia
ERR589743	58329126	ERS537278	ERP006678	NA	NA	China	Asia
ERR589744	60812544	ERS537279	ERP006678	NA	NA	China	Asia
ERR589745	73422402	ERS537280	ERP006678	NA	NA	China	Asia
ERR589746	67228824	ERS537281	ERP006678	NA	NA	China	Asia
ERR589747	76799894	ERS537282	ERP006678	NA	NA	China	Asia
ERR589748	65285702	ERS537283	ERP006678	NA	NA	China	Asia
ERR589749	66274242	ERS537284	ERP006678	NA	NA	China	Asia
ERR589750	51170028	ERS537285	ERP006678	NA	NA	China	Asia
ERR589751	57326354	ERS537286	ERP006678	NA	NA	China	Asia
ERR589752	56466820	ERS537287	ERP006678	NA	NA	China	Asia
ERR589753	61917212	ERS537288	ERP006678	NA	NA	China	Asia
ERR589754	54024340	ERS537289	ERP006678	NA	NA	China	Asia
ERR589755	60826178	ERS537290	ERP006678	NA	NA	China	Asia
ERR589756	81480756	ERS537291	ERP006678	NA	NA	China	Asia
ERR589757	70419588	ERS537292	ERP006678	NA	NA	China	Asia
ERR589758	71025578	ERS537293	ERP006678	NA	NA	China	Asia
ERR589759	66469000	ERS537294	ERP006678	NA	NA	China	Asia
ERR589760	68708254	ERS537295	ERP006678	NA	NA	China	Asia
ERR589761	60468232	ERS537296	ERP006678	NA	NA	China	Asia
ERR589762	61603504	ERS537297	ERP006678	NA	NA	China	Asia
ERR589763	67141942	ERS537298	ERP006678	NA	NA	China	Asia
ERR589764	64001188	ERS537299	ERP006678	NA	NA	China	Asia
ERR589765	61423104	ERS537300	ERP006678	NA	NA	China	Asia
ERR589766	58035924	ERS537301	ERP006678	NA	NA	China	Asia
ERR589767	59879368	ERS537302	ERP006678	NA	NA	China	Asia
ERR589768	75638300	ERS537303	ERP006678	NA	NA	China	Asia
ERR589769	84300186	ERS537304	ERP006678	NA	NA	China	Asia
ERR589770	74319090	ERS537305	ERP006678	NA	NA	China	Asia
ERR589771	59402220	ERS537306	ERP006678	NA	NA	China	Asia
ERR589772	63743994	ERS537307	ERP006678	NA	NA	China	Asia
ERR589773	65981634	ERS537308	ERP006678	NA	NA	China	Asia
ERR589774	68756164	ERS537309	ERP006678	NA	NA	China	Asia
ERR589775	56107304	ERS537310	ERP006678	NA	NA	China	Asia
ERR589776	59480792	ERS537311	ERP006678	NA	NA	China	Asia
ERR589777	58374382	ERS537312	ERP006678	NA	NA	China	Asia
ERR589778	60728828	ERS537313	ERP006678	NA	NA	China	Asia
ERR589779	77917418	ERS537314	ERP006678	NA	NA	China	Asia
ERR589780	82999424	ERS537315	ERP006678	NA	NA	China	Asia
ERR589781	62801144	ERS537316	ERP006678	NA	NA	China	Asia
ERR589782	81222934	ERS537317	ERP006678	NA	NA	China	Asia
ERR589783	58851980	ERS537318	ERP006678	NA	NA	China	Asia
ERR589784	68546926	ERS537319	ERP006678	NA	NA	China	Asia
ERR589785	70626620	ERS537320	ERP006678	NA	NA	China	Asia
ERR589789	63354380	ERS537324	ERP006678	NA	NA	China	Asia
ERR589790	60615168	ERS537325	ERP006678	NA	NA	China	Asia
ERR589791	61653490	ERS537326	ERP006678	NA	NA	China	Asia
ERR589792	60334418	ERS537327	ERP006678	NA	NA	China	Asia
ERR589793	59121378	ERS537328	ERP006678	NA	NA	China	Asia
ERR589794	59509144	ERS537329	ERP006678	NA	NA	China	Asia
ERR589795	60582764	ERS537330	ERP006678	NA	NA	China	Asia
ERR589796	64787204	ERS537331	ERP006678	NA	NA	China	Asia
ERR589797	58951134	ERS537332	ERP006678	NA	NA	China	Asia
ERR589798	51070894	ERS537333	ERP006678	NA	NA	China	Asia
ERR589799	58207982	ERS537334	ERP006678	NA	NA	China	Asia
ERR589800	59193066	ERS537335	ERP006678	NA	NA	China	Asia
ERR589801	64153550	ERS537336	ERP006678	NA	NA	China	Asia

ERR589802	55927116	ERS537337	ERP006678	NA	NA	China	Asia
ERR589803	79574344	ERS537338	ERP006678	NA	NA	China	Asia
ERR589804	77039500	ERS537339	ERP006678	NA	NA	China	Asia
ERR589805	70602322	ERS537340	ERP006678	NA	NA	China	Asia
ERR589806	78262078	ERS537341	ERP006678	NA	NA	China	Asia
ERR589807	66058264	ERS537342	ERP006678	NA	NA	China	Asia
ERR589808	65714338	ERS537343	ERP006678	NA	NA	China	Asia
ERR589809	78028262	ERS537344	ERP006678	NA	NA	China	Asia
ERR589810	69992296	ERS537345	ERP006678	NA	NA	China	Asia
ERR589811	74906904	ERS537346	ERP006678	NA	NA	China	Asia
ERR589812	68014646	ERS537347	ERP006678	NA	NA	China	Asia
ERR589813	61405790	ERS537348	ERP006678	NA	NA	China	Asia
ERR589814	63278596	ERS537349	ERP006678	NA	NA	China	Asia
ERR589815	66312984	ERS537350	ERP006678	NA	NA	China	Asia
ERR589816	74226068	ERS537351	ERP006678	NA	NA	China	Asia
ERR589817	73557292	ERS537352	ERP006678	NA	NA	China	Asia
ERR589818	62392402	ERS537353	ERP006678	NA	NA	China	Asia
ERR589819	57548758	ERS537354	ERP006678	NA	NA	China	Asia
ERR589820	71482694	ERS537355	ERP006678	NA	NA	China	Asia
ERR589821	60341574	ERS537356	ERP006678	NA	NA	China	Asia
ERR589823	68888566	ERS537358	ERP006678	NA	NA	China	Asia
ERR589824	119912950	ERS537359	ERP006678	NA	NA	China	Asia
ERR589825	52694728	ERS537360	ERP006678	NA	NA	China	Asia
ERR589826	74097828	ERS537361	ERP006678	NA	NA	China	Asia
ERR589827	80593220	ERS537362	ERP006678	NA	NA	China	Asia
ERR589828	75118348	ERS537363	ERP006678	NA	NA	China	Asia
ERR589829	53102296	ERS537364	ERP006678	NA	NA	China	Asia
ERR589830	71662278	ERS537365	ERP006678	NA	NA	China	Asia
ERR589831	70249688	ERS537366	ERP006678	NA	NA	China	Asia
ERR589832	65807188	ERS537367	ERP006678	NA	NA	China	Asia
ERR589833	78708166	ERS537368	ERP006678	NA	NA	China	Asia
ERR589834	77260992	ERS537369	ERP006678	NA	NA	China	Asia
ERR589835	84344218	ERS537370	ERP006678	NA	NA	China	Asia
ERR589837	55908578	ERS537372	ERP006678	NA	NA	China	Asia
ERR589838	56854292	ERS537373	ERP006678	NA	NA	China	Asia
ERR589839	77485390	ERS537374	ERP006678	NA	NA	China	Asia
ERR589840	70375996	ERS537375	ERP006678	NA	NA	China	Asia
ERR589841	73484640	ERS537376	ERP006678	NA	NA	China	Asia
ERR589842	71951834	ERS537377	ERP006678	NA	NA	China	Asia
ERR589843	61413880	ERS537378	ERP006678	NA	NA	China	Asia
ERR589844	55112894	ERS537379	ERP006678	NA	NA	China	Asia
ERR589845	57457414	ERS537380	ERP006678	NA	NA	China	Asia
ERR589846	62762364	ERS537381	ERP006678	NA	NA	China	Asia
ERR589847	79315704	ERS537382	ERP006678	NA	NA	China	Asia
ERR589848	71622760	ERS537383	ERP006678	NA	NA	China	Asia
ERR589849	73697930	ERS537384	ERP006678	NA	NA	China	Asia
ERR589850	75233606	ERS537385	ERP006678	NA	NA	China	Asia
ERR589851	63367448	ERS537386	ERP006678	NA	NA	China	Asia
ERR589852	88633956	ERS537387	ERP006678	NA	NA	China	Asia
ERR589853	75014500	ERS537388	ERP006678	NA	NA	China	Asia
ERR589854	73931946	ERS537389	ERP006678	NA	NA	China	Asia
ERR589855	78993296	ERS537390	ERP006678	NA	NA	China	Asia
ERR589856	76008160	ERS537391	ERP006678	NA	NA	China	Asia
ERR589857	68961210	ERS537392	ERP006678	NA	NA	China	Asia
ERR589858	69819088	ERS537393	ERP006678	NA	NA	China	Asia
ERR589859	82500834	ERS537394	ERP006678	NA	NA	China	Asia
ERR589860	67693160	ERS537395	ERP006678	NA	NA	China	Asia
ERR589861	70400922	ERS537396	ERP006678	NA	NA	China	Asia
ERR589862	54715374	ERS537397	ERP006678	NA	NA	China	Asia
ERR589863	63743852	ERS537398	ERP006678	NA	NA	China	Asia
ERR589864	78998530	ERS537399	ERP006678	NA	NA	China	Asia
ERR589865	60521174	ERS537400	ERP006678	NA	NA	China	Asia
ERR589870	62871430	ERS537405	ERP006678	NA	NA	China	Asia
ERR589871	65281292	ERS537406	ERP006678	NA	NA	China	Asia
ERR589872	69335846	ERS537407	ERP006678	NA	NA	China	Asia
ERR589873	58512214	ERS537408	ERP006678	NA	NA	China	Asia
ERR589875	61293380	ERS537410	ERP006678	NA	NA	China	Asia
ERR636349	147984098	ERS554192	ERP007090	Healthy	NA	Sweden	Europe
ERR636351	113377860	ERS554194	ERP007090	Healthy	Urban	Sweden	Europe
ERR636352	184603796	ERS554195	ERP007090	Healthy	NA	Sweden	Europe
ERR636353	224437174	ERS554196	ERP007090	Healthy	Urban	Sweden	Europe
ERR636354	206293040	ERS554197	ERP007090	Healthy	NA	Sweden	Europe

ERR636355	201010060	ERS554198	ERP007090	Healthy	NA	Sweden	Europe
ERR636356	160774976	ERS554199	ERP007090	Healthy	NA	Sweden	Europe
ERR636357	194849272	ERS554200	ERP007090	Healthy	NA	Sweden	Europe
ERR636360	150106114	ERS554203	ERP007090	Healthy	NA	Sweden	Europe
ERR636361	233971282	ERS554204	ERP007090	Healthy	NA	Sweden	Europe
ERR636362	129734170	ERS554205	ERP007090	Healthy	NA	Sweden	Europe
ERR636363	195145886	ERS554206	ERP007090	Healthy	NA	Sweden	Europe
ERR636364	189041654	ERS554207	ERP007090	Healthy	NA	Sweden	Europe
ERR636365	173451850	ERS554208	ERP007090	Healthy	NA	Sweden	Europe
ERR636366	328342060	ERS554209	ERP007090	Healthy	NA	Sweden	Europe
ERR636367	240987672	ERS554210	ERP007090	Healthy	NA	Sweden	Europe
ERR636369	182408648	ERS554212	ERP007090	Healthy	NA	Sweden	Europe
ERR636370	255430132	ERS554213	ERP007090	Healthy	NA	Sweden	Europe
ERR636371	188188370	ERS554214	ERP007090	Healthy	NA	Sweden	Europe
ERR636372	155502422	ERS554215	ERP007090	Healthy	NA	Sweden	Europe
ERR636373	155592346	ERS554216	ERP007090	Healthy	NA	Sweden	Europe
ERR636374	279539812	ERS554217	ERP007090	Healthy	NA	Sweden	Europe
ERR636375	356009214	ERS554218	ERP007090	Healthy	NA	Sweden	Europe
ERR636376	109975282	ERS554219	ERP007090	Healthy	NA	Sweden	Europe
ERR636377	222913068	ERS554220	ERP007090	Healthy	NA	Sweden	Europe
ERR636378	205684656	ERS554221	ERP007090	Healthy	NA	Sweden	Europe
ERR636379	204235228	ERS554222	ERP007090	Healthy	NA	Sweden	Europe
ERR636380	137333440	ERS554223	ERP007090	Healthy	NA	Sweden	Europe
ERR636381	197507338	ERS554224	ERP007090	Healthy	NA	Sweden	Europe
ERR636382	151740678	ERS554225	ERP007090	Healthy	NA	Sweden	Europe
ERR636383	141934656	ERS554226	ERP007090	Healthy	NA	Sweden	Europe
ERR636384	170271930	ERS554227	ERP007090	Healthy	NA	Sweden	Europe
ERR636385	205724324	ERS554228	ERP007090	Healthy	NA	Sweden	Europe
ERR636386	161675760	ERS554229	ERP007090	Healthy	NA	Sweden	Europe
ERR636387	258076764	ERS554230	ERP007090	Healthy	NA	Sweden	Europe
ERR636388	209350240	ERS554231	ERP007090	Healthy	NA	Sweden	Europe
ERR636389	199797130	ERS554232	ERP007090	Healthy	NA	Sweden	Europe
ERR636390	155602906	ERS554233	ERP007090	Healthy	NA	Sweden	Europe
ERR636392	190599596	ERS554235	ERP007090	Healthy	NA	Sweden	Europe
ERR636393	234713530	ERS554236	ERP007090	Healthy	NA	Sweden	Europe
ERR636394	222875304	ERS554237	ERP007090	Healthy	NA	Sweden	Europe
ERR636395	214195050	ERS554238	ERP007090	Healthy	NA	Sweden	Europe
ERR636396	279021086	ERS554239	ERP007090	Healthy	NA	Sweden	Europe
ERR636397	150534486	ERS554240	ERP007090	Healthy	NA	Sweden	Europe
ERR636398	251408826	ERS554241	ERP007090	Healthy	NA	Sweden	Europe
ERR636399	182053822	ERS554242	ERP007090	Healthy	NA	Sweden	Europe
ERR636400	103411914	ERS554243	ERP007090	Healthy	NA	Sweden	Europe
ERR636401	337157882	ERS554244	ERP007090	Healthy	NA	Sweden	Europe
ERR636402	163268324	ERS554245	ERP007090	Healthy	NA	Sweden	Europe
ERR636403	221683292	ERS554246	ERP007090	Healthy	NA	Sweden	Europe
ERR636404	149499716	ERS554247	ERP007090	Healthy	NA	Sweden	Europe
ERR636405	169536970	ERS554248	ERP007090	Healthy	NA	Sweden	Europe
ERR636406	211975946	ERS554249	ERP007090	Healthy	NA	Sweden	Europe
ERR636407	192676610	ERS554250	ERP007090	Healthy	NA	Sweden	Europe
ERR636408	211307660	ERS554251	ERP007090	Healthy	NA	Sweden	Europe
ERR636409	222599268	ERS554252	ERP007090	Healthy	NA	Sweden	Europe
ERR636410	178808848	ERS554253	ERP007090	Healthy	NA	Sweden	Europe
ERR636411	173419070	ERS554254	ERP007090	Healthy	NA	Sweden	Europe
ERR636412	66072614	ERS554255	ERP007090	Healthy	NA	Sweden	Europe
ERR636413	53597772	ERS554256	ERP007090	Healthy	NA	Sweden	Europe
ERR636414	239483092	ERS554257	ERP007090	Healthy	NA	Sweden	Europe
ERR636415	160578082	ERS554258	ERP007090	Healthy	NA	Sweden	Europe
ERR636416	248995670	ERS554259	ERP007090	Healthy	NA	Sweden	Europe
ERR636417	221068190	ERS554260	ERP007090	Healthy	NA	Sweden	Europe
ERR636418	286857862	ERS554261	ERP007090	Healthy	NA	Sweden	Europe
ERR688506	64949772	ERS608499	ERP008729	Healthy	Urban	Austria	Europe
ERR688507	59678576	ERS608489	ERP008729	Healthy	Urban	Austria	Europe
ERR688509	51026634	ERS608521	ERP008729	Healthy	Urban	Austria	Europe
ERR688512	64219446	ERS608563	ERP008729	Diseased	Urban	Austria	Europe
ERR688513	52765100	ERS608620	ERP008729	Diseased	Urban	Austria	Europe
ERR688515	51674954	ERS608602	ERP008729	Diseased	Urban	Austria	Europe
ERR688517	51032080	ERS608589	ERP008729	Diseased	Urban	Austria	Europe
ERR688519	60704266	ERS608565	ERP008729	Diseased	Urban	Austria	Europe
ERR688521	53235370	ERS608598	ERP008729	Diseased	Urban	Austria	Europe
ERR688522	54400176	ERS608544	ERP008729	Diseased	Urban	Austria	Europe
ERR688523	50768166	ERS608527	ERP008729	Healthy	Urban	Austria	Europe
ERR688526	56841496	ERS608555	ERP008729	Diseased	Urban	Austria	Europe

ERR688527	59854508	ERS608605	ERP008729	Diseased	Urban	Austria	Europe
ERR688528	57206702	ERS608495	ERP008729	Healthy	Urban	Austria	Europe
ERR688529	57709000	ERS608505	ERP008729	Healthy	Urban	Austria	Europe
ERR688530	55960176	ERS608524	ERP008729	Healthy	Urban	Austria	Europe
ERR688533	52968250	ERS608515	ERP008729	Healthy	Urban	Austria	Europe
ERR688534	62616384	ERS608616	ERP008729	Diseased	Urban	Austria	Europe
ERR688535	62068202	ERS608507	ERP008729	Healthy	Urban	Austria	Europe
ERR688536	59869858	ERS608531	ERP008729	Diseased	Urban	Austria	Europe
ERR688537	58730900	ERS608501	ERP008729	Healthy	Urban	Austria	Europe
ERR688538	66934520	ERS608542	ERP008729	Diseased	Urban	Austria	Europe
ERR688539	58615384	ERS608477	ERP008729	Healthy	Urban	Austria	Europe
ERR688541	63160316	ERS608575	ERP008729	Diseased	Urban	Austria	Europe
ERR688542	66161212	ERS608557	ERP008729	Diseased	Urban	Austria	Europe
ERR688544	51482232	ERS608496	ERP008729	Healthy	Urban	Austria	Europe
ERR688546	58400498	ERS608536	ERP008729	Diseased	Urban	Austria	Europe
ERR688548	58726772	ERS608617	ERP008729	Diseased	Urban	Austria	Europe
ERR688549	78341350	ERS608543	ERP008729	Diseased	Urban	Austria	Europe
ERR688553	55773544	ERS608579	ERP008729	Diseased	Urban	Austria	Europe
ERR688554	54288710	ERS608486	ERP008729	Healthy	Urban	Austria	Europe
ERR688555	50517366	ERS608567	ERP008729	Diseased	Urban	Austria	Europe
ERR688559	65637426	ERS608483	ERP008729	Healthy	Urban	Austria	Europe
ERR688560	58381130	ERS608573	ERP008729	Diseased	Urban	Austria	Europe
ERR688561	51098750	ERS608528	ERP008729	Healthy	Urban	Austria	Europe
ERR688562	58199812	ERS608484	ERP008729	Healthy	Urban	Austria	Europe
ERR688564	59573400	ERS608503	ERP008729	Healthy	Urban	Austria	Europe
ERR688565	54331716	ERS608509	ERP008729	Healthy	Urban	Austria	Europe
ERR688569	62445124	ERS608577	ERP008729	Diseased	Urban	Austria	Europe
ERR688571	58720772	ERS608609	ERP008729	Diseased	Urban	Austria	Europe
ERR688572	64862254	ERS608601	ERP008729	Diseased	Urban	Austria	Europe
ERR688573	55343590	ERS608593	ERP008729	Diseased	Urban	Austria	Europe
ERR688574	55822274	ERS608611	ERP008729	Diseased	Urban	Austria	Europe
ERR688575	60455004	ERS608578	ERP008729	Diseased	Urban	Austria	Europe
ERR688576	59983982	ERS608603	ERP008729	Diseased	Urban	Austria	Europe
ERR688577	57372900	ERS608591	ERP008729	Diseased	Urban	Austria	Europe
ERR688578	54123104	ERS608582	ERP008729	Diseased	Urban	Austria	Europe
ERR688579	50341098	ERS608580	ERP008729	Diseased	Urban	Austria	Europe
ERR688580	54833492	ERS608618	ERP008729	Diseased	Urban	Austria	Europe
ERR688581	61435876	ERS608612	ERP008729	Diseased	Urban	Austria	Europe
ERR688582	66402356	ERS608576	ERP008729	Diseased	Urban	Austria	Europe
ERR688583	72405312	ERS608614	ERP008729	Diseased	Urban	Austria	Europe
ERR688586	50793540	ERS608556	ERP008729	Diseased	Urban	Austria	Europe
ERR688587	57807612	ERS608538	ERP008729	Diseased	Urban	Austria	Europe
ERR688589	60792560	ERS608572	ERP008729	Diseased	Urban	Austria	Europe
ERR688592	62864826	ERS608549	ERP008729	Diseased	Urban	Austria	Europe
ERR688596	56180634	ERS608502	ERP008729	Healthy	Urban	Austria	Europe
ERR688597	59668950	ERS608546	ERP008729	Diseased	Urban	Austria	Europe
ERR688598	56741188	ERS608487	ERP008729	Healthy	Urban	Austria	Europe
ERR688599	53250334	ERS608506	ERP008729	Healthy	Urban	Austria	Europe
ERR688601	54623602	ERS608533	ERP008729	Diseased	Urban	Austria	Europe
ERR688604	52291010	ERS608534	ERP008729	Diseased	Urban	Austria	Europe
ERR688606	55572074	ERS608491	ERP008729	Healthy	Urban	Austria	Europe
ERR688610	57108240	ERS608493	ERP008729	Healthy	Urban	Austria	Europe
ERR688611	55282336	ERS608545	ERP008729	Diseased	Urban	Austria	Europe
ERR688613	61860240	ERS608570	ERP008729	Diseased	Urban	Austria	Europe
ERR688615	54619932	ERS608571	ERP008729	Diseased	Urban	Austria	Europe
ERR688618	50738108	ERS608583	ERP008729	Diseased	Urban	Austria	Europe
ERR688619	65447322	ERS608492	ERP008729	Healthy	Urban	Austria	Europe
ERR688620	58216314	ERS608566	ERP008729	Diseased	Urban	Austria	Europe
ERR688621	58832068	ERS608537	ERP008729	Diseased	Urban	Austria	Europe
ERR688623	52753574	ERS608568	ERP008729	Diseased	Urban	Austria	Europe
ERR688624	61330000	ERS608510	ERP008729	Healthy	Urban	Austria	Europe
ERR688625	57245288	ERS608548	ERP008729	Diseased	Urban	Austria	Europe
ERR688626	56479204	ERS608535	ERP008729	Diseased	Urban	Austria	Europe
ERR688627	50006782	ERS608569	ERP008729	Diseased	Urban	Austria	Europe
ERR688628	62578104	ERS608539	ERP008729	Diseased	Urban	Austria	Europe
ERR688630	50866300	ERS608550	ERP008729	Diseased	Urban	Austria	Europe
ERR688632	60168106	ERS608574	ERP008729	Diseased	Urban	Austria	Europe
ERR688633	65420990	ERS608551	ERP008729	Diseased	Urban	Austria	Europe
ERR688634	58034758	ERS608597	ERP008729	Diseased	Urban	Austria	Europe
ERR688636	56240176	ERS608584	ERP008729	Diseased	Urban	Austria	Europe
ERR688637	60201910	ERS608608	ERP008729	Diseased	Urban	Austria	Europe
ERR688638	73716782	ERS608619	ERP008729	Diseased	Urban	Austria	Europe

ERR688639	51839146	ERS608600	ERP008729	Diseased	Urban	Austria	Europe
ERR688640	57303106	ERS608581	ERP008729	Diseased	Urban	Austria	Europe
ERR688641	56934344	ERS608595	ERP008729	Diseased	Urban	Austria	Europe
ERR688642	63608370	ERS608615	ERP008729	Diseased	Urban	Austria	Europe
ERR688645	63804608	ERS608532	ERP008729	Diseased	Urban	Austria	Europe
ERR688646	53758206	ERS608520	ERP008729	Healthy	Urban	Austria	Europe
ERR688647	53425978	ERS608592	ERP008729	Diseased	Urban	Austria	Europe
ERR688650	66030386	ERS608613	ERP008729	Diseased	Urban	Austria	Europe
ERR710427	54255502	ERS631827	ERP008729	Healthy	Urban	Austria	Europe
ERR710428	50267724	ERS631828	ERP008729	Healthy	Urban	Austria	Europe
ERR710432	51100774	ERS631833	ERP008729	Healthy	Urban	Austria	Europe
ERR911953	64334768	ERS746710	ERP010700	Healthy	Urban	United Kingdom	Europe
ERR911954	101742444	ERS746711	ERP010700	Healthy	NA	United Kingdom	Europe
ERR911955	80104466	ERS746712	ERP010700	Healthy	NA	United Kingdom	Europe
ERR911956	77741276	ERS746713	ERP010700	Healthy	NA	United Kingdom	Europe
ERR911957	80235494	ERS746714	ERP010700	Healthy	NA	United Kingdom	Europe
ERR911958	72383506	ERS746715	ERP010700	Healthy	Urban	United Kingdom	Europe
ERR911959	73863708	ERS746716	ERP010700	Healthy	Urban	United Kingdom	Europe
ERR911960	75026366	ERS746717	ERP010700	Healthy	Urban	United Kingdom	Europe
ERR911961	63950580	ERS746718	ERP010700	Healthy	Urban	United Kingdom	Europe
ERR911962	69838136	ERS746719	ERP010700	Healthy	Urban	United Kingdom	Europe
ERR911963	76076414	ERS746720	ERP010700	Healthy	NA	United Kingdom	Europe
ERR911964	89621746	ERS746721	ERP010700	Healthy	NA	United Kingdom	Europe
ERR911965	58380666	ERS746722	ERP010700	Healthy	Urban	United Kingdom	Europe
ERR911966	87970740	ERS746723	ERP010700	Healthy	NA	United Kingdom	Europe
ERR911967	70809438	ERS746724	ERP010700	Healthy	Urban	United Kingdom	Europe
ERR911968	70955428	ERS746725	ERP010700	Healthy	NA	United Kingdom	Europe
ERR911969	63859746	ERS746726	ERP010700	Healthy	NA	United Kingdom	Europe
ERR911970	73685638	ERS746727	ERP010700	Healthy	NA	United Kingdom	Europe
ERR911971	63173506	ERS746728	ERP010700	Healthy	NA	United Kingdom	Europe
ERR911972	64297714	ERS746729	ERP010700	Healthy	NA	United Kingdom	Europe
ERR911973	68218004	ERS746730	ERP010700	Healthy	NA	United Kingdom	Europe
ERR911974	83021310	ERS746731	ERP010700	Healthy	NA	United Kingdom	Europe
ERR911975	69371420	ERS746732	ERP010700	Healthy	NA	United Kingdom	Europe
ERR911976	64762690	ERS746733	ERP010700	Healthy	NA	United Kingdom	Europe
ERR911977	66991690	ERS746734	ERP010700	Healthy	NA	United Kingdom	Europe
ERR911978	72040492	ERS746735	ERP010700	Healthy	NA	United Kingdom	Europe
ERR911979	71322338	ERS746736	ERP010700	Healthy	NA	United Kingdom	Europe
ERR911980	70298090	ERS746737	ERP010700	Healthy	NA	United Kingdom	Europe
ERR911981	76985332	ERS746738	ERP010700	Healthy	NA	United Kingdom	Europe
ERR911982	71003370	ERS746739	ERP010700	Healthy	NA	United Kingdom	Europe
ERR911983	73660606	ERS746740	ERP010700	Healthy	NA	United Kingdom	Europe
ERR911984	61090142	ERS746741	ERP010700	Healthy	NA	United Kingdom	Europe
ERR911985	72956344	ERS746742	ERP010700	Healthy	NA	United Kingdom	Europe
ERR911986	57277118	ERS746743	ERP010700	Healthy	NA	United Kingdom	Europe
ERR911987	65705858	ERS746744	ERP010700	Healthy	NA	United Kingdom	Europe
ERR911988	68269436	ERS746745	ERP010700	Healthy	NA	United Kingdom	Europe
ERR911989	60875008	ERS746746	ERP010700	Healthy	NA	United Kingdom	Europe
ERR911990	76212070	ERS746747	ERP010700	Healthy	NA	United Kingdom	Europe
ERR911991	71384794	ERS746748	ERP010700	Healthy	NA	United Kingdom	Europe
ERR911992	69921792	ERS746749	ERP010700	Healthy	NA	United Kingdom	Europe
ERR911993	109919284	ERS746750	ERP010700	Healthy	NA	United Kingdom	Europe
ERR911994	59531308	ERS746751	ERP010700	Healthy	NA	United Kingdom	Europe
ERR911995	91536868	ERS746752	ERP010700	Healthy	NA	United Kingdom	Europe
ERR911996	75771848	ERS746753	ERP010700	Healthy	NA	United Kingdom	Europe
ERR911997	73162768	ERS746754	ERP010700	Healthy	NA	United Kingdom	Europe
ERR911998	81859770	ERS746755	ERP010700	Healthy	NA	United Kingdom	Europe
ERR912000	64998346	ERS746757	ERP010700	Healthy	NA	United Kingdom	Europe
ERR912001	67442488	ERS746758	ERP010700	Healthy	NA	United Kingdom	Europe
ERR912002	64536254	ERS746759	ERP010700	Healthy	NA	United Kingdom	Europe
ERR912003	67783670	ERS746760	ERP010700	Healthy	NA	United Kingdom	Europe
ERR912005	76916356	ERS746762	ERP010700	Healthy	NA	United Kingdom	Europe
ERR912006	78673440	ERS746763	ERP010700	Healthy	NA	United Kingdom	Europe
ERR912007	71912988	ERS746764	ERP010700	Healthy	NA	United Kingdom	Europe
ERR912008	70843278	ERS746765	ERP010700	Healthy	NA	United Kingdom	Europe
ERR912009	63539244	ERS746766	ERP010700	Healthy	NA	United Kingdom	Europe
ERR912010	60242432	ERS746767	ERP010700	Healthy	NA	United Kingdom	Europe
ERR912011	80183770	ERS746768	ERP010700	Healthy	NA	United Kingdom	Europe
ERR912012	71747436	ERS746769	ERP010700	Healthy	NA	United Kingdom	Europe
ERR912013	68939842	ERS746770	ERP010700	Healthy	NA	United Kingdom	Europe
ERR912014	62820030	ERS746771	ERP010700	Healthy	NA	United Kingdom	Europe
ERR912016	55144692	ERS746773	ERP010700	Healthy	NA	United Kingdom	Europe

ERR912017	67218384	ERS746774	ERP010700	Healthy	NA	United Kingdom	Europe
ERR912018	70013440	ERS746775	ERP010700	Healthy	NA	United Kingdom	Europe
ERR912019	67834872	ERS746776	ERP010700	Healthy	NA	United Kingdom	Europe
ERR912020	73790852	ERS746777	ERP010700	Healthy	NA	United Kingdom	Europe
ERR912021	77338230	ERS746778	ERP010700	Healthy	NA	United Kingdom	Europe
ERR912022	61427076	ERS746779	ERP010700	Healthy	NA	United Kingdom	Europe
ERR912023	62082972	ERS746780	ERP010700	Healthy	NA	United Kingdom	Europe
ERR912024	54487290	ERS746781	ERP010700	Healthy	NA	United Kingdom	Europe
ERR912025	90552120	ERS746782	ERP010700	Healthy	NA	United Kingdom	Europe
ERR912026	76045890	ERS746783	ERP010700	Healthy	NA	United Kingdom	Europe
ERR912027	69512328	ERS746784	ERP010700	Healthy	NA	United Kingdom	Europe
ERR912028	82796788	ERS746785	ERP010700	Healthy	NA	United Kingdom	Europe
ERR912029	60465110	ERS746786	ERP010700	Healthy	NA	United Kingdom	Europe
ERR912030	62567154	ERS746787	ERP010700	Healthy	NA	United Kingdom	Europe
ERR912031	67175178	ERS746788	ERP010700	Healthy	NA	United Kingdom	Europe
ERR912032	77251548	ERS746789	ERP010700	Healthy	NA	United Kingdom	Europe
ERR912033	73449520	ERS746790	ERP010700	Healthy	NA	United Kingdom	Europe
ERR912034	78689342	ERS746791	ERP010700	Healthy	NA	United Kingdom	Europe
ERR912035	67292782	ERS746792	ERP010700	Healthy	NA	United Kingdom	Europe
ERR912036	95908144	ERS746793	ERP010700	Healthy	NA	United Kingdom	Europe
ERR912037	74724306	ERS746794	ERP010700	Healthy	Urban	United Kingdom	Europe
ERR912038	75098186	ERS746795	ERP010700	Healthy	NA	United Kingdom	Europe
ERR912039	79055658	ERS746796	ERP010700	Healthy	NA	United Kingdom	Europe
ERR912040	53836552	ERS746797	ERP010700	Healthy	NA	United Kingdom	Europe
ERR912041	73686642	ERS746798	ERP010700	Healthy	NA	United Kingdom	Europe
ERR912042	79916932	ERS746799	ERP010700	Healthy	NA	United Kingdom	Europe
ERR912043	76016266	ERS746800	ERP010700	Healthy	NA	United Kingdom	Europe
ERR912044	70437356	ERS746801	ERP010700	Healthy	NA	United Kingdom	Europe
ERR912045	62827462	ERS746802	ERP010700	Healthy	NA	United Kingdom	Europe
ERR912046	83665358	ERS746803	ERP010700	Healthy	NA	United Kingdom	Europe
ERR912047	75368498	ERS746804	ERP010700	Healthy	NA	United Kingdom	Europe
ERR912048	79154778	ERS746805	ERP010700	Healthy	NA	United Kingdom	Europe
ERR912049	76802326	ERS746806	ERP010700	Healthy	NA	United Kingdom	Europe
ERR912050	80839326	ERS746807	ERP010700	Healthy	NA	United Kingdom	Europe
ERR912051	79122312	ERS746808	ERP010700	Healthy	NA	United Kingdom	Europe
ERR912052	77642648	ERS746809	ERP010700	Healthy	NA	United Kingdom	Europe
ERR912053	76515990	ERS746810	ERP010700	Healthy	NA	United Kingdom	Europe
ERR912054	76306922	ERS746811	ERP010700	Healthy	NA	United Kingdom	Europe
ERR912055	79007544	ERS746812	ERP010700	Healthy	NA	United Kingdom	Europe
ERR912056	60658346	ERS746813	ERP010700	Healthy	NA	United Kingdom	Europe
ERR912057	72512598	ERS746814	ERP010700	Healthy	NA	United Kingdom	Europe
ERR912058	83003516	ERS746815	ERP010700	Healthy	NA	United Kingdom	Europe
ERR912059	82965520	ERS746816	ERP010700	Healthy	NA	United Kingdom	Europe
ERR912060	64273822	ERS746817	ERP010700	Healthy	NA	United Kingdom	Europe
ERR912061	70285276	ERS746818	ERP010700	Healthy	NA	United Kingdom	Europe
ERR912062	83181462	ERS746819	ERP010700	Healthy	NA	United Kingdom	Europe
ERR912063	65416512	ERS746820	ERP010700	Healthy	NA	United Kingdom	Europe
ERR912064	77844198	ERS746821	ERP010700	Healthy	NA	United Kingdom	Europe
ERR912065	62455456	ERS746822	ERP010700	Healthy	NA	United Kingdom	Europe
ERR912066	100407626	ERS746823	ERP010700	Healthy	NA	United Kingdom	Europe
ERR912067	58704052	ERS746824	ERP010700	Healthy	NA	United Kingdom	Europe
ERR912068	76289936	ERS746825	ERP010700	Healthy	NA	United Kingdom	Europe
ERR912069	62844092	ERS746826	ERP010700	Healthy	NA	United Kingdom	Europe
ERR912070	69685338	ERS746827	ERP010700	Healthy	NA	United Kingdom	Europe
ERR912071	70512380	ERS746828	ERP010700	Healthy	NA	United Kingdom	Europe
ERR912072	77603044	ERS746829	ERP010700	Healthy	NA	United Kingdom	Europe
ERR912073	68291882	ERS746830	ERP010700	Healthy	NA	United Kingdom	Europe
ERR912074	75225000	ERS746831	ERP010700	Healthy	NA	United Kingdom	Europe
ERR912075	66478752	ERS746832	ERP010700	Healthy	NA	United Kingdom	Europe
ERR912076	80692314	ERS746833	ERP010700	Healthy	NA	United Kingdom	Europe
ERR912077	60786080	ERS746834	ERP010700	Healthy	NA	United Kingdom	Europe
ERR912078	69903008	ERS746835	ERP010700	Healthy	NA	United Kingdom	Europe
ERR912079	72963810	ERS746836	ERP010700	Healthy	NA	United Kingdom	Europe
ERR912080	67134470	ERS746837	ERP010700	Healthy	NA	United Kingdom	Europe
ERR912081	73127632	ERS746838	ERP010700	Healthy	NA	United Kingdom	Europe
ERR912082	68934032	ERS746839	ERP010700	Healthy	Urban	United Kingdom	Europe
ERR912083	70950898	ERS746840	ERP010700	Healthy	NA	United Kingdom	Europe
ERR912084	76631768	ERS746841	ERP010700	Healthy	NA	United Kingdom	Europe
ERR912085	81105752	ERS746842	ERP010700	Healthy	NA	United Kingdom	Europe
ERR912086	76105304	ERS746843	ERP010700	Healthy	NA	United Kingdom	Europe
ERR912087	100697702	ERS746844	ERP010700	Healthy	NA	United Kingdom	Europe
ERR912088	65737600	ERS746845	ERP010700	Healthy	NA	United Kingdom	Europe

ERR912089	75775524	ERS746846	ERP010700	Healthy	NA	United Kingdom	Europe
ERR912090	64996720	ERS746847	ERP010700	Healthy	NA	United Kingdom	Europe
ERR912091	59831200	ERS746848	ERP010700	Healthy	NA	United Kingdom	Europe
ERR912092	74692310	ERS746849	ERP010700	Healthy	NA	United Kingdom	Europe
ERR912093	61356204	ERS746850	ERP010700	Healthy	NA	United Kingdom	Europe
ERR912094	73480976	ERS746851	ERP010700	Healthy	NA	United Kingdom	Europe
ERR912095	73736870	ERS746852	ERP010700	Healthy	NA	United Kingdom	Europe
ERR912096	74832300	ERS746853	ERP010700	Healthy	NA	United Kingdom	Europe
ERR912097	85346590	ERS746854	ERP010700	Healthy	NA	United Kingdom	Europe
ERR912098	61469786	ERS746855	ERP010700	Healthy	NA	United Kingdom	Europe
ERR912099	80640908	ERS746856	ERP010700	Healthy	NA	United Kingdom	Europe
ERR912100	69609424	ERS746857	ERP010700	Healthy	NA	United Kingdom	Europe
ERR912101	82420002	ERS746858	ERP010700	Healthy	NA	United Kingdom	Europe
ERR912102	69159446	ERS746859	ERP010700	Healthy	NA	United Kingdom	Europe
ERR912103	60771458	ERS746860	ERP010700	Healthy	NA	United Kingdom	Europe
ERR912104	74526922	ERS746861	ERP010700	Healthy	NA	United Kingdom	Europe
ERR912105	73428274	ERS746862	ERP010700	Healthy	NA	United Kingdom	Europe
ERR912106	86000274	ERS746863	ERP010700	Healthy	NA	United Kingdom	Europe
ERR912107	74791338	ERS746864	ERP010700	Healthy	NA	United Kingdom	Europe
ERR912108	78597718	ERS746865	ERP010700	Healthy	NA	United Kingdom	Europe
ERR912109	69838878	ERS746866	ERP010700	Healthy	NA	United Kingdom	Europe
ERR912110	69959040	ERS746867	ERP010700	Healthy	NA	United Kingdom	Europe
ERR912111	65434240	ERS746868	ERP010700	Healthy	NA	United Kingdom	Europe
ERR912112	73398602	ERS746869	ERP010700	Healthy	NA	United Kingdom	Europe
ERR912113	68326712	ERS746870	ERP010700	Healthy	NA	United Kingdom	Europe
ERR912114	71720130	ERS746871	ERP010700	Healthy	NA	United Kingdom	Europe
ERR912115	74516056	ERS746872	ERP010700	Healthy	NA	United Kingdom	Europe
ERR912116	83487500	ERS746873	ERP010700	Healthy	NA	United Kingdom	Europe
ERR912117	83300828	ERS746874	ERP010700	Healthy	NA	United Kingdom	Europe
ERR912118	74338068	ERS746875	ERP010700	Healthy	NA	United Kingdom	Europe
ERR912119	59005502	ERS746876	ERP010700	Healthy	NA	United Kingdom	Europe
ERR912120	72568340	ERS746877	ERP010700	Healthy	NA	United Kingdom	Europe
ERR912121	67105912	ERS746878	ERP010700	Healthy	NA	United Kingdom	Europe
ERR912122	56909442	ERS746879	ERP010700	Healthy	NA	United Kingdom	Europe
ERR912123	80038942	ERS746880	ERP010700	Healthy	NA	United Kingdom	Europe
ERR912124	75150146	ERS746881	ERP010700	Healthy	NA	United Kingdom	Europe
ERR912125	73280940	ERS746882	ERP010700	Healthy	NA	United Kingdom	Europe
ERR912126	60473804	ERS746883	ERP010700	Healthy	NA	United Kingdom	Europe
ERR912127	73253462	ERS746884	ERP010700	Healthy	NA	United Kingdom	Europe
ERR912128	71246888	ERS746885	ERP010700	Healthy	NA	United Kingdom	Europe
ERR912129	65559736	ERS746886	ERP010700	Healthy	NA	United Kingdom	Europe
ERR912130	66987864	ERS746887	ERP010700	Healthy	NA	United Kingdom	Europe
ERR912131	73413078	ERS746888	ERP010700	Healthy	NA	United Kingdom	Europe
ERR912132	68749002	ERS746889	ERP010700	Healthy	NA	United Kingdom	Europe
ERR912133	73446786	ERS746890	ERP010700	Healthy	NA	United Kingdom	Europe
ERR912134	92350142	ERS746891	ERP010700	Healthy	NA	United Kingdom	Europe
ERR912135	70984064	ERS746892	ERP010700	Healthy	NA	United Kingdom	Europe
ERR912136	89618032	ERS746893	ERP010700	Healthy	NA	United Kingdom	Europe
ERR912137	75800840	ERS746894	ERP010700	Healthy	NA	United Kingdom	Europe
ERR912138	77699876	ERS746895	ERP010700	Healthy	NA	United Kingdom	Europe
ERR912139	87251408	ERS746896	ERP010700	Healthy	NA	United Kingdom	Europe
ERR912140	75068618	ERS746897	ERP010700	Healthy	NA	United Kingdom	Europe
ERR912141	66603632	ERS746898	ERP010700	Healthy	NA	United Kingdom	Europe
ERR912142	74145428	ERS746899	ERP010700	Healthy	NA	United Kingdom	Europe
ERR912143	68182836	ERS746900	ERP010700	Healthy	NA	United Kingdom	Europe
ERR912144	87288462	ERS746901	ERP010700	Healthy	NA	United Kingdom	Europe
ERR912145	70273270	ERS746902	ERP010700	Healthy	NA	United Kingdom	Europe
ERR912146	70529310	ERS746903	ERP010700	Healthy	NA	United Kingdom	Europe
ERR912147	75777254	ERS746904	ERP010700	Healthy	NA	United Kingdom	Europe
ERR912148	76680248	ERS746905	ERP010700	Healthy	NA	United Kingdom	Europe
ERR912149	83182342	ERS746906	ERP010700	Healthy	NA	United Kingdom	Europe
ERR912150	70670996	ERS746907	ERP010700	Healthy	NA	United Kingdom	Europe
ERR912151	75427966	ERS746908	ERP010700	Healthy	NA	United Kingdom	Europe
ERR912152	70586390	ERS746909	ERP010700	Healthy	NA	United Kingdom	Europe
ERR912153	83550276	ERS746910	ERP010700	Healthy	NA	United Kingdom	Europe
ERR912154	72024512	ERS746911	ERP010700	Healthy	NA	United Kingdom	Europe
ERR912155	80020142	ERS746912	ERP010700	Healthy	NA	United Kingdom	Europe
ERR912156	77615248	ERS746913	ERP010700	Healthy	NA	United Kingdom	Europe
ERR912157	64541890	ERS746914	ERP010700	Healthy	NA	United Kingdom	Europe
ERR912158	68294232	ERS746915	ERP010700	Healthy	NA	United Kingdom	Europe
ERR912159	71323536	ERS746916	ERP010700	Healthy	NA	United Kingdom	Europe
ERR912160	87724804	ERS746917	ERP010700	Healthy	NA	United Kingdom	Europe

ERR912161	76456844	ERS746918	ERP010700	Healthy	NA	United Kingdom	Europe
ERR912162	89197432	ERS746919	ERP010700	Healthy	NA	United Kingdom	Europe
ERR912163	82672166	ERS746920	ERP010700	Healthy	NA	United Kingdom	Europe
ERR912164	51948390	ERS746921	ERP010700	Healthy	NA	United Kingdom	Europe
ERR912165	75338046	ERS746922	ERP010700	Healthy	NA	United Kingdom	Europe
ERR912166	69703218	ERS746923	ERP010700	Healthy	NA	United Kingdom	Europe
ERR912167	67012002	ERS746924	ERP010700	Healthy	NA	United Kingdom	Europe
ERR912168	76286136	ERS746925	ERP010700	Healthy	NA	United Kingdom	Europe
ERR912169	65110806	ERS746926	ERP010700	Healthy	NA	United Kingdom	Europe
ERR912170	66633278	ERS746927	ERP010700	Healthy	NA	United Kingdom	Europe
ERR912171	63373492	ERS746928	ERP010700	Healthy	NA	United Kingdom	Europe
ERR912172	59965666	ERS746929	ERP010700	Healthy	NA	United Kingdom	Europe
ERR912173	78430516	ERS746930	ERP010700	Healthy	NA	United Kingdom	Europe
ERR912174	76952260	ERS746931	ERP010700	Healthy	NA	United Kingdom	Europe
ERR912175	68758972	ERS746932	ERP010700	Healthy	NA	United Kingdom	Europe
ERR912176	71971268	ERS746933	ERP010700	Healthy	NA	United Kingdom	Europe
ERR912177	84827394	ERS746934	ERP010700	Healthy	NA	United Kingdom	Europe
ERR912178	73150018	ERS746935	ERP010700	Healthy	NA	United Kingdom	Europe
ERR912179	71109022	ERS746936	ERP010700	Healthy	NA	United Kingdom	Europe
ERR912180	73824966	ERS746937	ERP010700	Healthy	NA	United Kingdom	Europe
ERR912181	73150164	ERS746938	ERP010700	Healthy	NA	United Kingdom	Europe
ERR912182	71251736	ERS746939	ERP010700	Healthy	NA	United Kingdom	Europe
ERR912183	91244818	ERS746940	ERP010700	Healthy	NA	United Kingdom	Europe
ERR912184	76771010	ERS746941	ERP010700	Healthy	NA	United Kingdom	Europe
ERR912185	68319378	ERS746942	ERP010700	Healthy	Urban	United Kingdom	Europe
ERR912186	73116452	ERS746943	ERP010700	Healthy	NA	United Kingdom	Europe
ERR912187	66708386	ERS746944	ERP010700	Healthy	NA	United Kingdom	Europe
ERR912188	63907412	ERS746945	ERP010700	Healthy	NA	United Kingdom	Europe
ERR912189	78031002	ERS746946	ERP010700	Healthy	NA	United Kingdom	Europe
ERR912190	65969320	ERS746947	ERP010700	Healthy	NA	United Kingdom	Europe
ERR912191	90083444	ERS746948	ERP010700	Healthy	NA	United Kingdom	Europe
ERR912192	63601526	ERS746949	ERP010700	Healthy	NA	United Kingdom	Europe
ERR912193	58800416	ERS746950	ERP010700	Healthy	NA	United Kingdom	Europe
ERR912194	75223688	ERS746951	ERP010700	Healthy	NA	United Kingdom	Europe
ERR912195	65648250	ERS746952	ERP010700	Healthy	NA	United Kingdom	Europe
ERR912196	59201358	ERS746953	ERP010700	Healthy	NA	United Kingdom	Europe
ERR912197	83208348	ERS746954	ERP010700	Healthy	NA	United Kingdom	Europe
ERR912198	86361636	ERS746955	ERP010700	Healthy	NA	United Kingdom	Europe
ERR912199	88450548	ERS746956	ERP010700	Healthy	NA	United Kingdom	Europe
ERR912200	81960192	ERS746957	ERP010700	Healthy	NA	United Kingdom	Europe
ERR912201	89020272	ERS746958	ERP010700	Healthy	NA	United Kingdom	Europe
ERR912202	63467504	ERS746959	ERP010700	Healthy	NA	United Kingdom	Europe
ERR1018185	60749206	ERS848618	ERP012177	Diseased	NA	China	Asia
ERR1018186	52854792	ERS848619	ERP012177	Diseased	NA	China	Asia
ERR1018187	59873382	ERS848620	ERP012177	Diseased	NA	China	Asia
ERR1018188	55761702	ERS848621	ERP012177	Diseased	NA	China	Asia
ERR1018189	64434830	ERS848622	ERP012177	Diseased	NA	China	Asia
ERR1018190	52102228	ERS848623	ERP012177	Diseased	NA	China	Asia
ERR1018191	62456910	ERS848624	ERP012177	Diseased	NA	China	Asia
ERR1018192	60073374	ERS848625	ERP012177	Diseased	NA	China	Asia
ERR1018193	61229038	ERS848626	ERP012177	Healthy	NA	China	Asia
ERR1018194	61472286	ERS848627	ERP012177	Diseased	NA	China	Asia
ERR1018195	64164216	ERS848628	ERP012177	Healthy	NA	China	Asia
ERR1018196	66150638	ERS848629	ERP012177	Diseased	NA	China	Asia
ERR1018197	60074902	ERS848630	ERP012177	Diseased	NA	China	Asia
ERR1018198	57351612	ERS848631	ERP012177	Diseased	NA	China	Asia
ERR1018200	58546892	ERS848633	ERP012177	Diseased	NA	China	Asia
ERR1018201	56584998	ERS848634	ERP012177	Diseased	NA	China	Asia
ERR1018202	51292156	ERS848635	ERP012177	Diseased	NA	China	Asia
ERR1018203	54116632	ERS848636	ERP012177	Diseased	NA	China	Asia
ERR1018204	56648482	ERS848637	ERP012177	Diseased	NA	China	Asia
ERR1018205	57035188	ERS848638	ERP012177	Healthy	NA	China	Asia
ERR1018206	62258080	ERS848639	ERP012177	Diseased	NA	China	Asia
ERR1018208	58822118	ERS848641	ERP012177	Diseased	NA	China	Asia
ERR1018209	56134340	ERS848642	ERP012177	Healthy	NA	China	Asia
ERR1018210	60820472	ERS848643	ERP012177	Diseased	NA	China	Asia
ERR1018211	61937738	ERS848644	ERP012177	Diseased	NA	China	Asia
ERR1018212	59665954	ERS848645	ERP012177	Diseased	NA	China	Asia
ERR1018213	57476278	ERS848646	ERP012177	Diseased	NA	China	Asia
ERR1018214	62349482	ERS848647	ERP012177	Diseased	NA	China	Asia
ERR1018215	71617186	ERS848648	ERP012177	Diseased	NA	China	Asia
ERR1018217	63330138	ERS848650	ERP012177	Healthy	NA	China	Asia

ERR1018218	62319068	ERS848651	ERP012177	Diseased	NA	China	Asia
ERR1018219	72370444	ERS848652	ERP012177	Diseased	NA	China	Asia
ERR1018220	68159682	ERS848653	ERP012177	Diseased	NA	China	Asia
ERR1018221	61187040	ERS848654	ERP012177	Diseased	NA	China	Asia
ERR1018222	57034986	ERS848655	ERP012177	Diseased	NA	China	Asia
ERR1018223	61334626	ERS848656	ERP012177	Diseased	NA	China	Asia
ERR1018224	50308904	ERS848657	ERP012177	Diseased	NA	China	Asia
ERR1018226	68438974	ERS848659	ERP012177	Diseased	NA	China	Asia
ERR1018227	72260226	ERS848660	ERP012177	Diseased	NA	China	Asia
ERR1018228	62569760	ERS848661	ERP012177	Diseased	NA	China	Asia
ERR1018229	58539128	ERS848662	ERP012177	Diseased	NA	China	Asia
ERR1018230	59984388	ERS848663	ERP012177	Diseased	NA	China	Asia
ERR1018231	53627618	ERS848664	ERP012177	Diseased	NA	China	Asia
ERR1018232	66064102	ERS848665	ERP012177	Healthy	NA	China	Asia
ERR1018233	70428992	ERS848666	ERP012177	Healthy	NA	China	Asia
ERR1018234	61617318	ERS848667	ERP012177	Healthy	NA	China	Asia
ERR1018235	68472136	ERS848668	ERP012177	Healthy	NA	China	Asia
ERR1018236	69534266	ERS848669	ERP012177	Healthy	NA	China	Asia
ERR1018237	68004090	ERS848670	ERP012177	Healthy	NA	China	Asia
ERR1018238	72077438	ERS848671	ERP012177	Healthy	NA	China	Asia
ERR1018239	59159772	ERS848672	ERP012177	Healthy	NA	China	Asia
ERR1018240	67478554	ERS848673	ERP012177	Healthy	NA	China	Asia
ERR1018241	67318902	ERS848674	ERP012177	Diseased	NA	China	Asia
ERR1018242	63843786	ERS848675	ERP012177	Healthy	NA	China	Asia
ERR1018243	63808150	ERS848676	ERP012177	Healthy	NA	China	Asia
ERR1018244	60520634	ERS848677	ERP012177	Healthy	NA	China	Asia
ERR1018245	52698148	ERS848678	ERP012177	Healthy	NA	China	Asia
ERR1018246	52788942	ERS848679	ERP012177	Healthy	NA	China	Asia
ERR1018247	57114396	ERS848680	ERP012177	Healthy	NA	China	Asia
ERR1018248	58591508	ERS848681	ERP012177	Healthy	NA	China	Asia
ERR1018251	55499678	ERS848684	ERP012177	Healthy	NA	China	Asia
ERR1018254	56707948	ERS848687	ERP012177	Healthy	NA	China	Asia
ERR1018257	61595372	ERS848690	ERP012177	Healthy	NA	China	Asia
ERR1018258	57962360	ERS848691	ERP012177	Healthy	NA	China	Asia
ERR1018259	56561906	ERS848692	ERP012177	Healthy	NA	China	Asia
ERR1018260	61955812	ERS848693	ERP012177	Healthy	NA	China	Asia
ERR1018261	54321366	ERS848694	ERP012177	Healthy	NA	China	Asia
ERR1018262	77263508	ERS848695	ERP012177	Healthy	NA	China	Asia
ERR1018263	55579348	ERS848696	ERP012177	Healthy	NA	China	Asia
ERR1018264	59038886	ERS848697	ERP012177	Healthy	NA	China	Asia
ERR1018265	62784088	ERS848698	ERP012177	Healthy	NA	China	Asia
ERR1018266	61975626	ERS848699	ERP012177	Healthy	NA	China	Asia
ERR1018267	51471774	ERS848700	ERP012177	Healthy	NA	China	Asia
ERR1018268	58364678	ERS848701	ERP012177	Healthy	NA	China	Asia
ERR1018269	60676490	ERS848702	ERP012177	Healthy	NA	China	Asia
ERR1018270	63341086	ERS848703	ERP012177	Healthy	NA	China	Asia
ERR1018271	55190022	ERS848704	ERP012177	Healthy	NA	China	Asia
ERR1018272	60252226	ERS848705	ERP012177	Healthy	NA	China	Asia
ERR1018273	62929376	ERS848706	ERP012177	Healthy	NA	China	Asia
ERR1018274	63280364	ERS848707	ERP012177	Healthy	NA	China	Asia
ERR1018277	60671012	ERS848710	ERP012177	Healthy	NA	China	Asia
ERR1018278	63322690	ERS848711	ERP012177	Healthy	NA	China	Asia
ERR1018279	63781394	ERS848712	ERP012177	Diseased	NA	China	Asia
ERR1018280	65133586	ERS848713	ERP012177	Diseased	NA	China	Asia
ERR1018281	52187750	ERS848714	ERP012177	Diseased	NA	China	Asia
ERR1018282	56903946	ERS848715	ERP012177	Diseased	NA	China	Asia
ERR1018283	51245744	ERS848716	ERP012177	Diseased	NA	China	Asia
ERR1018284	51812238	ERS848717	ERP012177	Diseased	NA	China	Asia
ERR1018285	62149784	ERS848718	ERP012177	Diseased	NA	China	Asia
ERR1018286	55720964	ERS848719	ERP012177	Diseased	NA	China	Asia
ERR1018299	50539146	ERS848732	ERP012177	Diseased	NA	China	Asia
ERR1018302	65450728	ERS848735	ERP012177	Diseased	NA	China	Asia
ERR1018303	57315876	ERS848736	ERP012177	Diseased	NA	China	Asia
ERR1018304	59042440	ERS848737	ERP012177	Diseased	NA	China	Asia
ERR1018305	58518194	ERS848738	ERP012177	Diseased	NA	China	Asia
ERR1018307	51070446	ERS848740	ERP012177	Diseased	NA	China	Asia
ERR1018308	58080274	ERS848741	ERP012177	Healthy	NA	China	Asia
ERR1018309	57219434	ERS848742	ERP012177	Healthy	NA	China	Asia
ERR1018310	66186116	ERS848743	ERP012177	Diseased	NA	China	Asia
ERR1018311	58789886	ERS848744	ERP012177	Diseased	NA	China	Asia
ERR1018312	63339956	ERS848745	ERP012177	Diseased	NA	China	Asia
ERR1190532	54716998	ERS1015607	ERP013562	NA	NA	China	Asia

ERR1190533	51609568	ERS1015608	ERP013562	NA	NA	China	Asia
ERR1190544	50746282	ERS1015619	ERP013562	NA	NA	China	Asia
ERR1190548	51713314	ERS1015623	ERP013562	NA	NA	China	Asia
ERR1190551	55289040	ERS1015626	ERP013562	NA	NA	China	Asia
ERR1190552	53844364	ERS1015627	ERP013562	NA	NA	China	Asia
ERR1190554	51827064	ERS1015629	ERP013562	NA	NA	China	Asia
ERR1190555	51558962	ERS1015630	ERP013562	NA	NA	China	Asia
ERR1190557	51158560	ERS1015632	ERP013562	NA	NA	China	Asia
ERR1190558	57144988	ERS1015633	ERP013562	NA	NA	China	Asia
ERR1190559	52828966	ERS1015634	ERP013562	NA	NA	China	Asia
ERR1190562	52344108	ERS1015637	ERP013562	NA	NA	China	Asia
ERR1190563	50816498	ERS1015638	ERP013562	NA	NA	China	Asia
ERR1190565	54960196	ERS1015640	ERP013562	NA	NA	China	Asia
ERR1190566	50575282	ERS1015641	ERP013562	NA	NA	China	Asia
ERR1190567	50194958	ERS1015642	ERP013562	NA	NA	China	Asia
ERR1190568	50815440	ERS1015643	ERP013562	NA	NA	China	Asia
ERR1190569	61370222	ERS1015644	ERP013562	NA	NA	China	Asia
ERR1190570	53292288	ERS1015645	ERP013562	NA	NA	China	Asia
ERR1190571	50193400	ERS1015646	ERP013562	NA	NA	China	Asia
ERR1190572	53223496	ERS1015647	ERP013562	NA	NA	China	Asia
ERR1190573	58737052	ERS1015648	ERP013562	NA	NA	China	Asia
ERR1190576	55885762	ERS1015651	ERP013562	NA	NA	China	Asia
ERR1190577	52643868	ERS1015652	ERP013562	NA	NA	China	Asia
ERR1190578	53561260	ERS1015653	ERP013562	NA	NA	China	Asia
ERR1190581	54444882	ERS1015656	ERP013562	NA	NA	China	Asia
ERR1190582	54392656	ERS1015657	ERP013562	NA	NA	China	Asia
ERR1190583	55427670	ERS1015658	ERP013562	NA	NA	China	Asia
ERR1190593	55270590	ERS1015668	ERP013562	NA	NA	China	Asia
ERR1190594	52837240	ERS1015669	ERP013562	NA	NA	China	Asia
ERR1190595	52133154	ERS1015670	ERP013562	NA	NA	China	Asia
ERR1190596	61296272	ERS1015671	ERP013562	NA	NA	China	Asia
ERR1190597	53926258	ERS1015672	ERP013562	NA	NA	China	Asia
ERR1190598	58396400	ERS1015673	ERP013562	NA	NA	China	Asia
ERR1190599	54817232	ERS1015674	ERP013562	NA	NA	China	Asia
ERR1190600	56595370	ERS1015675	ERP013562	NA	NA	China	Asia
ERR1190601	54393054	ERS1015676	ERP013562	NA	NA	China	Asia
ERR1190602	50827210	ERS1015677	ERP013562	NA	NA	China	Asia
ERR1190606	60531266	ERS1015681	ERP013562	NA	NA	China	Asia
ERR1190608	57053884	ERS1015683	ERP013562	NA	NA	China	Asia
ERR1190609	54004094	ERS1015684	ERP013562	NA	NA	China	Asia
ERR1190610	50116580	ERS1015685	ERP013562	NA	NA	China	Asia
ERR1190611	50923202	ERS1015686	ERP013562	NA	NA	China	Asia
ERR1190612	51595212	ERS1015687	ERP013562	NA	NA	China	Asia
ERR1190613	51375296	ERS1015688	ERP013562	NA	NA	China	Asia
ERR1190614	57621920	ERS1015689	ERP013562	NA	NA	China	Asia
ERR1190615	53825754	ERS1015690	ERP013562	NA	NA	China	Asia
ERR1190616	55178696	ERS1015691	ERP013562	NA	NA	China	Asia
ERR1190617	63760016	ERS1015692	ERP013562	NA	NA	China	Asia
ERR1190619	52422404	ERS1015694	ERP013562	NA	NA	China	Asia
ERR1190620	50993470	ERS1015695	ERP013562	NA	NA	China	Asia
ERR1190622	51347560	ERS1015697	ERP013562	NA	NA	China	Asia
ERR1190624	52112406	ERS1015699	ERP013562	NA	NA	China	Asia
ERR1190625	63242536	ERS1015700	ERP013562	NA	NA	China	Asia
ERR1190626	63238470	ERS1015701	ERP013562	NA	NA	China	Asia
ERR1190629	57277928	ERS1015704	ERP013562	NA	NA	China	Asia
ERR1190630	55331490	ERS1015705	ERP013562	NA	NA	China	Asia
ERR1190633	54816544	ERS1015708	ERP013562	NA	NA	China	Asia
ERR1190634	55436978	ERS1015709	ERP013562	NA	NA	China	Asia
ERR1190638	56011668	ERS1015713	ERP013562	NA	NA	China	Asia
ERR1190639	61523632	ERS1015714	ERP013562	NA	NA	China	Asia
ERR1190640	54767610	ERS1015715	ERP013562	NA	NA	China	Asia
ERR1190644	56036934	ERS1015719	ERP013562	NA	NA	China	Asia
ERR1190647	54623334	ERS1015722	ERP013562	NA	NA	China	Asia
ERR1190650	57361388	ERS1015725	ERP013562	NA	NA	China	Asia
ERR1190651	56324416	ERS1015726	ERP013562	NA	NA	China	Asia
ERR1190654	51463784	ERS1015729	ERP013562	NA	NA	China	Asia
ERR1190657	60129676	ERS1015732	ERP013562	NA	NA	China	Asia
ERR1190658	53382606	ERS1015733	ERP013562	NA	NA	China	Asia
ERR1190659	56584782	ERS1015734	ERP013562	NA	NA	China	Asia
ERR1190660	50975556	ERS1015735	ERP013562	NA	NA	China	Asia
ERR1190661	50265884	ERS1015736	ERP013562	NA	NA	China	Asia
ERR1190667	50726736	ERS1015742	ERP013562	NA	NA	China	Asia

ERR1190669	50118674	ERS1015744	ERP013562	NA	NA	China	Asia
ERR1190670	55920376	ERS1015745	ERP013562	NA	NA	China	Asia
ERR1190674	52744228	ERS1015749	ERP013562	NA	NA	China	Asia
ERR1190675	53285670	ERS1015750	ERP013562	NA	NA	China	Asia
ERR1190676	58414648	ERS1015751	ERP013562	NA	NA	China	Asia
ERR1190677	60219146	ERS1015752	ERP013562	NA	NA	China	Asia
ERR1190678	51328138	ERS1015753	ERP013562	NA	NA	China	Asia
ERR1190684	50318240	ERS1015759	ERP013562	NA	NA	China	Asia
ERR1190686	52301838	ERS1015761	ERP013562	NA	NA	China	Asia
ERR1190693	50188346	ERS1015768	ERP013562	NA	NA	China	Asia
ERR1190695	52798752	ERS1015770	ERP013562	NA	NA	China	Asia
ERR1190697	59634450	ERS1015772	ERP013562	NA	NA	China	Asia
ERR1190699	51960330	ERS1015774	ERP013562	NA	NA	China	Asia
ERR1190700	50677114	ERS1015775	ERP013562	NA	NA	China	Asia
ERR1190702	53118688	ERS1015777	ERP013562	NA	NA	China	Asia
ERR1190704	56222968	ERS1015779	ERP013562	NA	NA	China	Asia
ERR1190705	51321626	ERS1015780	ERP013562	NA	NA	China	Asia
ERR1190706	53272122	ERS1015781	ERP013562	NA	NA	China	Asia
ERR1190707	53258072	ERS1015782	ERP013562	NA	NA	China	Asia
ERR1190712	54492366	ERS1015787	ERP013562	NA	NA	China	Asia
ERR1190714	56679546	ERS1015789	ERP013562	NA	NA	China	Asia
ERR1190715	51832786	ERS1015790	ERP013562	NA	NA	China	Asia
ERR1190716	52869022	ERS1015791	ERP013562	NA	NA	China	Asia
ERR1190717	58597042	ERS1015792	ERP013562	NA	NA	China	Asia
ERR1190718	53994162	ERS1015793	ERP013562	NA	NA	China	Asia
ERR1190719	58169740	ERS1015794	ERP013562	NA	NA	China	Asia
ERR1190720	58959848	ERS1015795	ERP013562	NA	NA	China	Asia
ERR1190721	52550852	ERS1015796	ERP013562	NA	NA	China	Asia
ERR1190722	58109016	ERS1015797	ERP013562	NA	NA	China	Asia
ERR1190723	61646704	ERS1015798	ERP013562	NA	NA	China	Asia
ERR1190724	57413768	ERS1015799	ERP013562	NA	NA	China	Asia
ERR1190726	54470982	ERS1015801	ERP013562	NA	NA	China	Asia
ERR1190727	55574998	ERS1015802	ERP013562	NA	NA	China	Asia
ERR1190728	50901386	ERS1015803	ERP013562	NA	NA	China	Asia
ERR1190730	51917946	ERS1015805	ERP013562	NA	NA	China	Asia
ERR1190734	53467454	ERS1015809	ERP013562	NA	NA	China	Asia
ERR1190735	52908444	ERS1015810	ERP013562	NA	NA	China	Asia
ERR1190736	54580542	ERS1015811	ERP013562	NA	NA	China	Asia
ERR1190737	53295482	ERS1015812	ERP013562	NA	NA	China	Asia
ERR1190738	58156520	ERS1015813	ERP013562	NA	NA	China	Asia
ERR1190739	51973120	ERS1015814	ERP013562	NA	NA	China	Asia
ERR1190740	55382612	ERS1015815	ERP013562	NA	NA	China	Asia
ERR1190741	52417516	ERS1015816	ERP013562	NA	NA	China	Asia
ERR1190742	58910316	ERS1015817	ERP013562	NA	NA	China	Asia
ERR1190743	51493194	ERS1015818	ERP013562	NA	NA	China	Asia
ERR1190744	64720706	ERS1015819	ERP013562	NA	NA	China	Asia
ERR1190745	67140266	ERS1015820	ERP013562	NA	NA	China	Asia
ERR1190746	57729814	ERS1015821	ERP013562	NA	NA	China	Asia
ERR1190748	51821088	ERS1015823	ERP013562	NA	NA	China	Asia
ERR1190749	50292850	ERS1015824	ERP013562	NA	NA	China	Asia
ERR1190753	53399622	ERS1015828	ERP013562	NA	NA	China	Asia
ERR1190754	55859714	ERS1015829	ERP013562	NA	NA	China	Asia
ERR1190755	53370020	ERS1015830	ERP013562	NA	NA	China	Asia
ERR1190756	55678778	ERS1015831	ERP013562	NA	NA	China	Asia
ERR1190757	50474292	ERS1015832	ERP013562	NA	NA	China	Asia
ERR1190758	59508186	ERS1015833	ERP013562	NA	NA	China	Asia
ERR1190759	64977588	ERS1015834	ERP013562	NA	NA	China	Asia
ERR1190760	63859734	ERS1015835	ERP013562	NA	NA	China	Asia
ERR1190761	67131192	ERS1015836	ERP013562	NA	NA	China	Asia
ERR1190762	53239870	ERS1015837	ERP013562	NA	NA	China	Asia
ERR1190763	55005486	ERS1015838	ERP013562	NA	NA	China	Asia
ERR1190764	57275164	ERS1015839	ERP013562	NA	NA	China	Asia
ERR1190765	64836184	ERS1015840	ERP013562	NA	NA	China	Asia
ERR1190766	66322084	ERS1015841	ERP013562	NA	NA	China	Asia
ERR1190767	56422164	ERS1015842	ERP013562	NA	NA	China	Asia
ERR1190768	53051560	ERS1015843	ERP013562	NA	NA	China	Asia
ERR1190769	65731280	ERS1015844	ERP013562	NA	NA	China	Asia
ERR1190770	86200096	ERS1015845	ERP013562	NA	NA	China	Asia
ERR1190771	72561012	ERS1015846	ERP013562	NA	NA	China	Asia
ERR1190772	79479230	ERS1015847	ERP013562	NA	NA	China	Asia
ERR1190773	79121106	ERS1015848	ERP013562	NA	NA	China	Asia
ERR1190774	78886648	ERS1015849	ERP013562	NA	NA	China	Asia

ERR1190775	69804040	ERS1015850	ERP013562	NA	NA	China	Asia
ERR1190776	52478606	ERS1015851	ERP013562	NA	NA	China	Asia
ERR1190777	71337932	ERS1015852	ERP013562	NA	NA	China	Asia
ERR1190778	73433330	ERS1015853	ERP013562	NA	NA	China	Asia
ERR1190779	61767682	ERS1015854	ERP013562	NA	NA	China	Asia
ERR1190780	71711180	ERS1015855	ERP013562	NA	NA	China	Asia
ERR1190781	76597608	ERS1015856	ERP013562	NA	NA	China	Asia
ERR1190782	79341658	ERS1015857	ERP013562	NA	NA	China	Asia
ERR1190783	77599644	ERS1015858	ERP013562	NA	NA	China	Asia
ERR1190784	79444992	ERS1015859	ERP013562	NA	NA	China	Asia
ERR1190785	77004772	ERS1015860	ERP013562	NA	NA	China	Asia
ERR1190786	69113904	ERS1015861	ERP013562	NA	NA	China	Asia
ERR1190787	56141134	ERS1015862	ERP013562	NA	NA	China	Asia
ERR1190788	71347376	ERS1015863	ERP013562	NA	NA	China	Asia
ERR1190789	116800962	ERS1015864	ERP013563	Diseased	NA	China	Asia
ERR1190790	106452446	ERS1015865	ERP013563	Diseased	NA	China	Asia
ERR1190792	60570378	ERS1015867	ERP013563	Diseased	NA	China	Asia
ERR1190793	105546390	ERS1015868	ERP013563	Diseased	NA	China	Asia
ERR1190794	105218768	ERS1015869	ERP013563	Diseased	NA	China	Asia
ERR1190796	102913454	ERS1015871	ERP013563	Diseased	NA	China	Asia
ERR1190797	108788416	ERS1015872	ERP013563	Diseased	NA	China	Asia
ERR1190798	137531378	ERS1015873	ERP013563	Diseased	NA	China	Asia
ERR1190799	89618526	ERS1015874	ERP013563	NA	NA	China	Asia
ERR1190800	120706360	ERS1015875	ERP013563	Diseased	NA	China	Asia
ERR1190801	97553410	ERS1015876	ERP013563	Diseased	NA	China	Asia
ERR1190802	139258258	ERS1015877	ERP013563	NA	NA	China	Asia
ERR1190803	97246910	ERS1015878	ERP013563	NA	NA	China	Asia
ERR1190804	89902694	ERS1015879	ERP013563	NA	NA	China	Asia
ERR1190805	102522976	ERS1015880	ERP013563	Diseased	NA	China	Asia
ERR1190806	108247610	ERS1015881	ERP013563	Diseased	NA	China	Asia
ERR1190807	96073702	ERS1015882	ERP013563	Diseased	NA	China	Asia
ERR1190808	125259768	ERS1015883	ERP013563	NA	NA	China	Asia
ERR1190809	123692256	ERS1015884	ERP013563	Diseased	NA	China	Asia
ERR1190810	98756358	ERS1015885	ERP013563	Diseased	NA	China	Asia
ERR1190811	60703426	ERS1015886	ERP013563	Diseased	NA	China	Asia
ERR1190812	98529996	ERS1015887	ERP013563	Diseased	NA	China	Asia
ERR1190813	119112562	ERS1015888	ERP013563	NA	NA	China	Asia
ERR1190814	113384384	ERS1015889	ERP013563	Diseased	NA	China	Asia
ERR1190815	118102740	ERS1015890	ERP013563	Diseased	NA	China	Asia
ERR1190816	114158134	ERS1015891	ERP013563	Diseased	NA	China	Asia
ERR1190817	54125590	ERS1015892	ERP013563	Diseased	NA	China	Asia
ERR1190818	57888172	ERS1015893	ERP013563	Diseased	NA	China	Asia
ERR1190819	56363348	ERS1015894	ERP013563	Diseased	NA	China	Asia
ERR1190820	57455562	ERS1015895	ERP013563	Diseased	NA	China	Asia
ERR1190821	92373634	ERS1015896	ERP013563	Diseased	NA	China	Asia
ERR1190822	114174864	ERS1015897	ERP013563	Diseased	NA	China	Asia
ERR1190823	126142380	ERS1015898	ERP013563	Diseased	NA	China	Asia
ERR1190824	98639454	ERS1015899	ERP013563	Diseased	NA	China	Asia
ERR1190826	56305970	ERS1015901	ERP013563	Diseased	NA	China	Asia
ERR1190827	99683724	ERS1015902	ERP013563	Diseased	NA	China	Asia
ERR1190828	119490276	ERS1015903	ERP013563	Diseased	NA	China	Asia
ERR1190829	50685670	ERS1015904	ERP013563	Diseased	NA	China	Asia
ERR1190830	52921694	ERS1015905	ERP013563	Diseased	NA	China	Asia
ERR1190831	55005368	ERS1015906	ERP013563	Diseased	NA	China	Asia
ERR1190832	52127740	ERS1015907	ERP013563	Diseased	NA	China	Asia
ERR1190833	54700110	ERS1015908	ERP013563	Diseased	NA	China	Asia
ERR1190834	52106340	ERS1015909	ERP013563	Diseased	NA	China	Asia
ERR1190835	50577500	ERS1015910	ERP013563	Diseased	NA	China	Asia
ERR1190836	56541436	ERS1015911	ERP013563	Diseased	NA	China	Asia
ERR1190837	55509950	ERS1015912	ERP013563	Diseased	NA	China	Asia
ERR1190838	52004160	ERS1015913	ERP013563	Diseased	NA	China	Asia
ERR1190839	51862584	ERS1015914	ERP013563	Diseased	NA	China	Asia
ERR1190840	57522048	ERS1015915	ERP013563	Diseased	NA	China	Asia
ERR1190842	60059602	ERS1015917	ERP013563	Diseased	NA	China	Asia
ERR1190843	59049384	ERS1015918	ERP013563	Diseased	NA	China	Asia
ERR1190845	54903210	ERS1015920	ERP013563	Diseased	NA	China	Asia
ERR1190846	57545752	ERS1015921	ERP013563	Diseased	NA	China	Asia
ERR1190847	51514284	ERS1015922	ERP013563	Diseased	NA	China	Asia
ERR1190848	52954024	ERS1015923	ERP013563	Diseased	NA	China	Asia
ERR1190849	50079620	ERS1015924	ERP013563	Diseased	NA	China	Asia
ERR1190850	50695222	ERS1015925	ERP013563	Diseased	NA	China	Asia
ERR1190851	54695738	ERS1015926	ERP013563	Diseased	NA	China	Asia

ERR1190852	55805060	ERS1015927	ERP013563	Diseased	NA	China	Asia
ERR1190853	59081002	ERS1015928	ERP013563	Diseased	NA	China	Asia
ERR1190854	51278508	ERS1015929	ERP013563	Diseased	NA	China	Asia
ERR1190855	87353516	ERS1015930	ERP013563	Diseased	NA	China	Asia
ERR1190856	54120098	ERS1015931	ERP013563	Diseased	NA	China	Asia
ERR1190858	53718052	ERS1015933	ERP013563	Diseased	NA	China	Asia
ERR1190859	60016894	ERS1015934	ERP013563	Diseased	NA	China	Asia
ERR1190860	54870888	ERS1015935	ERP013563	Diseased	NA	China	Asia
ERR1190861	50798650	ERS1015936	ERP013563	Diseased	NA	China	Asia
ERR1190862	51731082	ERS1015937	ERP013563	Diseased	NA	China	Asia
ERR1190863	51534634	ERS1015938	ERP013563	Diseased	NA	China	Asia
ERR1190866	70340604	ERS1015941	ERP013563	Diseased	NA	China	Asia
ERR1190867	59473148	ERS1015942	ERP013563	NA	NA	China	Asia
ERR1190868	61504248	ERS1015943	ERP013563	Diseased	NA	China	Asia
ERR1190869	56429078	ERS1015944	ERP013563	Diseased	NA	China	Asia
ERR1190870	50861562	ERS1015945	ERP013563	Diseased	NA	China	Asia
ERR1190871	112913978	ERS1015946	ERP013563	NA	NA	China	Asia
ERR1190872	121490350	ERS1015947	ERP013563	NA	NA	China	Asia
ERR1190873	50578320	ERS1015948	ERP013563	Diseased	NA	China	Asia
ERR1190874	51541964	ERS1015949	ERP013563	Diseased	NA	China	Asia
ERR1190875	127851016	ERS1015950	ERP013563	Diseased	NA	China	Asia
ERR1190876	57858390	ERS1015951	ERP013563	NA	NA	China	Asia
ERR1190878	62065556	ERS1015953	ERP013563	Diseased	NA	China	Asia
ERR1190879	109837962	ERS1015954	ERP013563	NA	NA	China	Asia
ERR1190880	98635342	ERS1015955	ERP013563	NA	NA	China	Asia
ERR1190881	136159766	ERS1015956	ERP013563	Diseased	NA	China	Asia
ERR1190882	110385444	ERS1015957	ERP013563	Diseased	NA	China	Asia
ERR1190883	108561410	ERS1015958	ERP013563	NA	NA	China	Asia
ERR1190884	100570944	ERS1015959	ERP013563	Diseased	NA	China	Asia
ERR1190885	156034130	ERS1015960	ERP013563	Diseased	NA	China	Asia
ERR1190886	116983264	ERS1015961	ERP013563	Diseased	NA	China	Asia
ERR1190887	99593678	ERS1015962	ERP013563	Diseased	NA	China	Asia
ERR1190888	105746162	ERS1015963	ERP013563	Diseased	NA	China	Asia
ERR1190890	70160496	ERS1015965	ERP013563	Diseased	NA	China	Asia
ERR1190891	108603906	ERS1015966	ERP013563	Diseased	NA	China	Asia
ERR1190892	103923104	ERS1015967	ERP013563	Diseased	NA	China	Asia
ERR1190893	59733716	ERS1015968	ERP013563	Diseased	NA	China	Asia
ERR1190895	130397090	ERS1015970	ERP013563	Diseased	NA	China	Asia
ERR1190896	105495408	ERS1015971	ERP013563	Diseased	NA	China	Asia
ERR1190897	107055678	ERS1015972	ERP013563	Diseased	NA	China	Asia
ERR1190898	85892376	ERS1015973	ERP013563	Diseased	NA	China	Asia
ERR1190899	117904942	ERS1015974	ERP013563	Diseased	NA	China	Asia
ERR1190900	123235084	ERS1015975	ERP013563	Diseased	NA	China	Asia
ERR1190901	147987406	ERS1015976	ERP013563	Diseased	NA	China	Asia
ERR1190902	115940686	ERS1015977	ERP013563	Diseased	NA	China	Asia
ERR1190903	101983440	ERS1015978	ERP013563	Diseased	NA	China	Asia
ERR1190904	116929376	ERS1015979	ERP013563	Diseased	NA	China	Asia
ERR1190905	124660714	ERS1015980	ERP013563	Diseased	NA	China	Asia
ERR1190906	84202806	ERS1015981	ERP013563	Diseased	NA	China	Asia
ERR1190907	63556518	ERS1015982	ERP013563	Diseased	NA	China	Asia
ERR1190908	55998200	ERS1015983	ERP013563	Diseased	NA	China	Asia
ERR1190910	50044138	ERS1015985	ERP013563	Diseased	NA	China	Asia
ERR1190911	116895270	ERS1015986	ERP013563	Diseased	NA	China	Asia
ERR1190912	119940274	ERS1015987	ERP013563	Diseased	NA	China	Asia
ERR1190913	126675908	ERS1015988	ERP013563	Diseased	NA	China	Asia
ERR1190914	109892944	ERS1015989	ERP013563	Diseased	NA	China	Asia
ERR1190915	55334938	ERS1015990	ERP013563	Diseased	NA	China	Asia
ERR1190916	54477628	ERS1015991	ERP013563	Diseased	NA	China	Asia
ERR1190917	55990680	ERS1015992	ERP013563	Diseased	NA	China	Asia
ERR1190918	51726572	ERS1015993	ERP013563	Diseased	NA	China	Asia
ERR1190920	64800362	ERS1015995	ERP013563	Diseased	NA	China	Asia
ERR1190921	52302694	ERS1015996	ERP013563	Diseased	NA	China	Asia
ERR1190922	61698266	ERS1015997	ERP013563	Diseased	NA	China	Asia
ERR1190923	52432584	ERS1015998	ERP013563	Diseased	NA	China	Asia
ERR1190924	122006484	ERS1015999	ERP013563	Diseased	NA	China	Asia
ERR1190925	54064682	ERS1016000	ERP013563	Diseased	NA	China	Asia
ERR1190926	89232606	ERS1016001	ERP013563	Diseased	NA	China	Asia
ERR1190927	56082774	ERS1016002	ERP013563	Diseased	NA	China	Asia
ERR1190928	51022428	ERS1016003	ERP013563	Diseased	NA	China	Asia
ERR1190929	51807856	ERS1016004	ERP013563	Diseased	NA	China	Asia
ERR1190930	85990970	ERS1016005	ERP013563	Diseased	NA	China	Asia
ERR1190931	60960806	ERS1016006	ERP013563	Diseased	NA	China	Asia

ERR1190932	62733310	ERS1016007	ERP013563	Diseased	NA	China	Asia
ERR1190933	56094862	ERS1016008	ERP013563	Diseased	NA	China	Asia
ERR1190934	57067308	ERS1016009	ERP013563	Diseased	NA	China	Asia
ERR1190936	112388158	ERS1016011	ERP013563	Diseased	NA	China	Asia
ERR1190937	52081824	ERS1016012	ERP013563	Diseased	NA	China	Asia
ERR1190938	50442370	ERS1016013	ERP013563	Diseased	NA	China	Asia
ERR1190939	57832868	ERS1016014	ERP013563	Diseased	NA	China	Asia
ERR1190940	100566804	ERS1016015	ERP013563	Diseased	NA	China	Asia
ERR1190941	65873810	ERS1016016	ERP013563	Diseased	NA	China	Asia
ERR1190942	107459074	ERS1016017	ERP013563	Diseased	NA	China	Asia
ERR1190943	50653500	ERS1016018	ERP013563	Diseased	NA	China	Asia
ERR1190944	115037502	ERS1016019	ERP013563	Diseased	NA	China	Asia
ERR1190945	56948136	ERS1016020	ERP013563	Diseased	NA	China	Asia
ERR1190946	124964372	ERS1016021	ERP013563	Diseased	NA	China	Asia
ERR1190947	52685080	ERS1016022	ERP013563	Diseased	NA	China	Asia
ERR1190948	57144328	ERS1016023	ERP013563	Diseased	NA	China	Asia
ERR1190949	50593588	ERS1016024	ERP013563	Diseased	NA	China	Asia
ERR1190952	55449936	ERS1016027	ERP013563	Diseased	NA	China	Asia
ERR1190953	50643352	ERS1016028	ERP013563	Diseased	NA	China	Asia
ERR1190954	52551370	ERS1016029	ERP013563	Diseased	NA	China	Asia
ERR1190956	53983810	ERS1016031	ERP013563	Diseased	NA	China	Asia
ERR1190957	52388774	ERS1016032	ERP013563	Diseased	NA	China	Asia
ERR1190959	56054548	ERS1016034	ERP013563	Diseased	NA	China	Asia
ERR1190960	54512394	ERS1016035	ERP013563	Diseased	NA	China	Asia
ERR1190961	50755700	ERS1016036	ERP013563	Diseased	NA	China	Asia
ERR1190962	54162204	ERS1016037	ERP013563	Diseased	NA	China	Asia
ERR1190964	52734012	ERS1016039	ERP013563	Diseased	NA	China	Asia
ERR1190965	58630310	ERS1016040	ERP013563	Diseased	NA	China	Asia
ERR1190966	55887086	ERS1016041	ERP013563	Diseased	NA	China	Asia
ERR1190967	55590046	ERS1016042	ERP013563	Diseased	NA	China	Asia
ERR1190969	52377710	ERS1016044	ERP013563	Diseased	NA	China	Asia
ERR1190971	52937364	ERS1016046	ERP013563	Diseased	NA	China	Asia
ERR1190972	53692224	ERS1016047	ERP013563	Diseased	NA	China	Asia
ERR1190974	60667570	ERS1016049	ERP013563	Diseased	NA	China	Asia
ERR1190976	53928460	ERS1016051	ERP013563	Diseased	NA	China	Asia
ERR1305877	96737010	ERS1076034	ERP014480	Diseased	Urban	Denmark	Europe
ERR1305878	66468770	ERS1076041	ERP014480	Diseased	Urban	Denmark	Europe
ERR1305879	92759304	ERS1076036	ERP014480	Diseased	Urban	Denmark	Europe
ERR1305880	65463296	ERS1076031	ERP014480	Diseased	Urban	Denmark	Europe
ERR1305881	72460578	ERS1076040	ERP014480	Diseased	Urban	Denmark	Europe
ERR1305882	64252258	ERS1076032	ERP014480	Diseased	Urban	Denmark	Europe
ERR1305883	69871514	ERS1076033	ERP014480	Diseased	Urban	Denmark	Europe
ERR1305884	78985070	ERS1076028	ERP014480	Diseased	Urban	Denmark	Europe
ERR1305885	71936520	ERS1076027	ERP014480	Diseased	Urban	Denmark	Europe
ERR1305886	59014798	ERS1076035	ERP014480	Diseased	Urban	Denmark	Europe
ERR1305887	79138072	ERS1076057	ERP014480	Diseased	Urban	Denmark	Europe
ERR1305888	71214170	ERS1076048	ERP014480	Diseased	Urban	Denmark	Europe
ERR1305889	71875782	ERS1076052	ERP014480	Diseased	Urban	Denmark	Europe
ERR1305890	68040932	ERS1076053	ERP014480	Diseased	Urban	Denmark	Europe
ERR1305891	80706110	ERS1076030	ERP014480	Diseased	Urban	Denmark	Europe
ERR1305892	102492224	ERS1076056	ERP014480	Diseased	Urban	Denmark	Europe
ERR1305893	74052272	ERS1076045	ERP014480	Diseased	Urban	Denmark	Europe
ERR1305894	62820488	ERS1076043	ERP014480	Diseased	Urban	Denmark	Europe
ERR1305895	76130996	ERS1076046	ERP014480	Diseased	Urban	Denmark	Europe
ERR1305896	67644044	ERS1076038	ERP014480	Diseased	Urban	Denmark	Europe
ERR1305897	79576766	ERS1076042	ERP014480	Diseased	Urban	Denmark	Europe
ERR1305899	69869278	ERS1076050	ERP014480	Diseased	Urban	Denmark	Europe
ERR1305900	63812190	ERS1076059	ERP014480	Diseased	Urban	Denmark	Europe
ERR1305901	86054730	ERS1076051	ERP014480	Diseased	Urban	Denmark	Europe
ERR1305902	72285260	ERS1076055	ERP014480	Diseased	Urban	Denmark	Europe
ERR1305903	57553992	ERS1076058	ERP014480	Diseased	Urban	Denmark	Europe
ERR1305904	82083404	ERS1076039	ERP014480	Diseased	Urban	Denmark	Europe
ERR1305905	76986918	ERS1076044	ERP014480	Diseased	Urban	Denmark	Europe
ERR1305906	74455024	ERS1076047	ERP014480	Diseased	Urban	Denmark	Europe
ERR1305907	72380414	ERS1076029	ERP014480	Diseased	Urban	Denmark	Europe
ERR1305908	78599378	ERS1076049	ERP014480	Diseased	Urban	Denmark	Europe
ERR1305909	76941010	ERS1076037	ERP014480	Diseased	Urban	Denmark	Europe
ERR1578619	63063792	ERS1289677	ERP016813	NA	NA	China	Asia
ERR1578620	62427590	ERS1289678	ERP016813	NA	NA	China	Asia
ERR1578621	56047472	ERS1289679	ERP016813	NA	NA	China	Asia
ERR1578622	68746102	ERS1289680	ERP016813	NA	NA	China	Asia
ERR1578623	52984672	ERS1289681	ERP016813	NA	NA	China	Asia

ERR1578624	55311318	ERS1289682	ERP016813	NA	NA	China	Asia
ERR1578625	56179694	ERS1289683	ERP016813	NA	NA	China	Asia
ERR1578626	62766668	ERS1289684	ERP016813	NA	NA	China	Asia
ERR1578627	58757372	ERS1289685	ERP016813	NA	NA	China	Asia
ERR1578628	61708680	ERS1289686	ERP016813	NA	NA	China	Asia
ERR1578629	59915634	ERS1289687	ERP016813	NA	NA	China	Asia
ERR1578630	59313788	ERS1289688	ERP016813	NA	NA	China	Asia
ERR1578631	60678538	ERS1289689	ERP016813	NA	NA	China	Asia
ERR1578632	55113826	ERS1289690	ERP016813	NA	NA	China	Asia
ERR1578633	55983568	ERS1289691	ERP016813	NA	NA	China	Asia
ERR1578634	57413220	ERS1289692	ERP016813	NA	NA	China	Asia
ERR1578635	60410492	ERS1289693	ERP016813	NA	NA	China	Asia
ERR1578636	61935242	ERS1289694	ERP016813	NA	NA	China	Asia
ERR1578637	50284238	ERS1289695	ERP016813	NA	NA	China	Asia
ERR1578638	61436002	ERS1289696	ERP016813	NA	NA	China	Asia
ERR1578639	71489108	ERS1289697	ERP016813	NA	NA	China	Asia
ERR1578640	69771298	ERS1289698	ERP016813	NA	NA	China	Asia
ERR1578643	57700040	ERS1289701	ERP016813	NA	NA	China	Asia
ERR1578645	61890862	ERS1289703	ERP016813	NA	NA	China	Asia
ERR1578646	67412122	ERS1289704	ERP016813	NA	NA	China	Asia
ERR1578647	60036972	ERS1289705	ERP016813	NA	NA	China	Asia
ERR1578648	61242478	ERS1289706	ERP016813	NA	NA	China	Asia
ERR1578649	60875178	ERS1289707	ERP016813	NA	NA	China	Asia
ERR1578650	68781928	ERS1289708	ERP016813	NA	NA	China	Asia
ERR1578651	57514190	ERS1289709	ERP016813	NA	NA	China	Asia
ERR1578652	62872406	ERS1289710	ERP016813	NA	NA	China	Asia
ERR1578653	54617984	ERS1289711	ERP016813	NA	NA	China	Asia
ERR1578654	56631746	ERS1289712	ERP016813	NA	NA	China	Asia
ERR1578655	51157526	ERS1289713	ERP016813	NA	NA	China	Asia
ERR1578656	55901804	ERS1289714	ERP016813	NA	NA	China	Asia
ERR1578657	65609998	ERS1289715	ERP016813	NA	NA	China	Asia
ERR1578658	57025786	ERS1289716	ERP016813	NA	NA	China	Asia
ERR1578659	68802936	ERS1289717	ERP016813	NA	NA	China	Asia
ERR1578660	59458180	ERS1289718	ERP016813	NA	NA	China	Asia
ERR1578661	54211582	ERS1289719	ERP016813	NA	NA	China	Asia
ERR1578662	50418004	ERS1289720	ERP016813	NA	NA	China	Asia
ERR1578663	62765856	ERS1289721	ERP016813	NA	NA	China	Asia
ERR1578664	52462816	ERS1289722	ERP016813	NA	NA	China	Asia
ERR1578665	67031750	ERS1289723	ERP016813	NA	NA	China	Asia
ERR1578666	70100472	ERS1289724	ERP016813	NA	NA	China	Asia
ERR1578667	68536366	ERS1289725	ERP016813	NA	NA	China	Asia
ERR1578668	56760540	ERS1289726	ERP016813	NA	NA	China	Asia
ERR1578669	58983338	ERS1289727	ERP016813	NA	NA	China	Asia
ERR1578670	62354288	ERS1289728	ERP016813	NA	NA	China	Asia
ERR1578671	60823848	ERS1289729	ERP016813	NA	NA	China	Asia
ERR1578672	51827482	ERS1289730	ERP016813	NA	NA	China	Asia
ERR1578673	51529352	ERS1289731	ERP016813	NA	NA	China	Asia
ERR1578674	57376772	ERS1289732	ERP016813	NA	NA	China	Asia
ERR1578675	58763632	ERS1289733	ERP016813	NA	NA	China	Asia
ERR1578676	58568046	ERS1289734	ERP016813	NA	NA	China	Asia
ERR1578678	50396708	ERS1289736	ERP016813	NA	NA	China	Asia
ERR1578679	50895468	ERS1289737	ERP016813	NA	NA	China	Asia
ERR1578680	55190606	ERS1289738	ERP016813	NA	NA	China	Asia
ERR1578681	55454628	ERS1289739	ERP016813	NA	NA	China	Asia
ERR1578682	77603700	ERS1289740	ERP016813	NA	NA	China	Asia
ERR1578683	60024078	ERS1289741	ERP016813	NA	NA	China	Asia
ERR1578684	58538708	ERS1289742	ERP016813	NA	NA	China	Asia
ERR1578685	55536104	ERS1289743	ERP016813	NA	NA	China	Asia
ERR1578686	54529938	ERS1289744	ERP016813	NA	NA	China	Asia
ERR1578687	54917578	ERS1289745	ERP016813	NA	NA	China	Asia
ERR1578688	55473708	ERS1289746	ERP016813	NA	NA	China	Asia
ERR1578689	54123892	ERS1289747	ERP016813	NA	NA	China	Asia
ERR1578690	56348656	ERS1289748	ERP016813	NA	NA	China	Asia
ERR1578692	78013816	ERS1289750	ERP016813	NA	NA	China	Asia
ERR1578693	56435138	ERS1289751	ERP016813	NA	NA	China	Asia
ERR1578694	56936070	ERS1289752	ERP016813	NA	NA	China	Asia
ERR1578696	56525812	ERS1289754	ERP016813	NA	NA	China	Asia
ERR1578697	67970718	ERS1289755	ERP016813	NA	NA	China	Asia
ERR1578698	66309336	ERS1289756	ERP016813	NA	NA	China	Asia
ERR1578699	51520054	ERS1289757	ERP016813	NA	NA	China	Asia
ERR1578700	56142236	ERS1289758	ERP016813	NA	NA	China	Asia
ERR1578701	57830178	ERS1289759	ERP016813	NA	NA	China	Asia

ERR1578702	54213138	ERS1289760	ERP016813	NA	NA	China	Asia
ERR1578703	63020722	ERS1289761	ERP016813	NA	NA	China	Asia
ERR1578704	65137684	ERS1289762	ERP016813	NA	NA	China	Asia
ERR1578705	53100568	ERS1289763	ERP016813	NA	NA	China	Asia
ERR1578706	55209246	ERS1289764	ERP016813	NA	NA	China	Asia
ERR1578707	63930826	ERS1289765	ERP016813	NA	NA	China	Asia
ERR1578708	59026166	ERS1289766	ERP016813	NA	NA	China	Asia
ERR1578710	54114600	ERS1289768	ERP016813	NA	NA	China	Asia
ERR1578713	62809596	ERS1289771	ERP016813	NA	NA	China	Asia
ERR1578714	56266672	ERS1289772	ERP016813	NA	NA	China	Asia
ERR1578715	52767764	ERS1289773	ERP016813	NA	NA	China	Asia
ERR1578716	60647136	ERS1289774	ERP016813	NA	NA	China	Asia
ERR1578719	57259788	ERS1289777	ERP016813	NA	NA	China	Asia
ERR1578720	60907826	ERS1289778	ERP016813	NA	NA	China	Asia
ERR1620257	50497300	ERS1343319	ERP017091	Diseased	NA	China	Asia
ERR1620259	50002694	ERS1343321	ERP017091	Diseased	NA	China	Asia
ERR1620260	66644380	ERS1343322	ERP017091	Diseased	NA	China	Asia
ERR1620262	61862596	ERS1343324	ERP017091	Diseased	NA	China	Asia
ERR1620263	58359536	ERS1343325	ERP017091	Diseased	NA	China	Asia
ERR1620264	54536954	ERS1343326	ERP017091	Diseased	NA	China	Asia
ERR1620265	53693190	ERS1343327	ERP017091	Diseased	NA	China	Asia
ERR1620266	60360372	ERS1343328	ERP017091	Diseased	NA	China	Asia
ERR1620267	58387072	ERS1343329	ERP017091	Diseased	NA	China	Asia
ERR1620268	62789624	ERS1343330	ERP017091	Diseased	NA	China	Asia
ERR1620269	75952732	ERS1343331	ERP017091	Diseased	NA	China	Asia
ERR1620271	53224572	ERS1343333	ERP017091	Diseased	NA	China	Asia
ERR1620272	51560494	ERS1343334	ERP017091	Diseased	NA	China	Asia
ERR1620275	53381286	ERS1343337	ERP017091	Diseased	NA	China	Asia
ERR1620276	55338394	ERS1343338	ERP017091	Diseased	NA	China	Asia
ERR1620278	61628228	ERS1343340	ERP017091	Diseased	NA	China	Asia
ERR1620280	60520636	ERS1343342	ERP017091	Diseased	NA	China	Asia
ERR1620281	50001208	ERS1343343	ERP017091	Diseased	NA	China	Asia
ERR1620282	65239512	ERS1343344	ERP017091	Diseased	NA	China	Asia
ERR1620283	56264950	ERS1343345	ERP017091	Diseased	NA	China	Asia
ERR1620285	60263364	ERS1343347	ERP017091	Diseased	NA	China	Asia
ERR1620286	61731870	ERS1343348	ERP017091	Diseased	NA	China	Asia
ERR1620287	50692424	ERS1343349	ERP017091	Diseased	NA	China	Asia
ERR1620288	77586944	ERS1343350	ERP017091	Diseased	NA	China	Asia
ERR1620289	60134532	ERS1343351	ERP017091	Diseased	NA	China	Asia
ERR1620290	65649778	ERS1343352	ERP017091	Diseased	NA	China	Asia
ERR1620291	53766286	ERS1343353	ERP017091	Diseased	NA	China	Asia
ERR1620293	58569662	ERS1343355	ERP017091	Diseased	NA	China	Asia
ERR1620294	58421202	ERS1343356	ERP017091	Diseased	NA	China	Asia
ERR1620295	55334516	ERS1343357	ERP017091	Diseased	NA	China	Asia
ERR1620296	71520120	ERS1343358	ERP017091	Diseased	NA	China	Asia
ERR1620297	65337008	ERS1343359	ERP017091	Diseased	NA	China	Asia
ERR1620298	61830890	ERS1343360	ERP017091	Diseased	NA	China	Asia
ERR1620300	60349606	ERS1343362	ERP017091	Diseased	NA	China	Asia
ERR1620301	51989932	ERS1343363	ERP017091	Diseased	NA	China	Asia
ERR1620302	58093994	ERS1343364	ERP017091	Diseased	NA	China	Asia
ERR1620305	51712072	ERS1343367	ERP017091	Diseased	NA	China	Asia
ERR1620306	50114988	ERS1343368	ERP017091	Diseased	NA	China	Asia
ERR1620313	55647706	ERS1343375	ERP017091	Diseased	NA	China	Asia
ERR1620314	55829782	ERS1343376	ERP017091	Diseased	NA	China	Asia
ERR1620317	61434160	ERS1343379	ERP017091	Diseased	NA	China	Asia
ERR1620319	51902612	ERS1343381	ERP017091	Diseased	NA	China	Asia
ERR1620322	51330772	ERS1343384	ERP017091	Healthy	NA	China	Asia
ERR1620323	53364388	ERS1343385	ERP017091	Healthy	NA	China	Asia
ERR1620324	63889880	ERS1343386	ERP017091	Healthy	NA	China	Asia
ERR1620325	56322628	ERS1343387	ERP017091	Healthy	NA	China	Asia
ERR1620326	65936594	ERS1343388	ERP017091	Healthy	NA	China	Asia
ERR1620327	53226398	ERS1343389	ERP017091	Healthy	NA	China	Asia
ERR1620328	51431112	ERS1343390	ERP017091	Healthy	NA	China	Asia
ERR1620330	53362170	ERS1343392	ERP017091	Healthy	NA	China	Asia
ERR1620331	62815964	ERS1343393	ERP017091	Healthy	NA	China	Asia
ERR1620332	64961684	ERS1343394	ERP017091	Healthy	NA	China	Asia
ERR1620334	63303022	ERS1343396	ERP017091	Healthy	NA	China	Asia
ERR1620335	68079856	ERS1343397	ERP017091	Healthy	NA	China	Asia
ERR1620336	65056178	ERS1343398	ERP017091	Healthy	NA	China	Asia
ERR1620337	58950628	ERS1343399	ERP017091	Healthy	NA	China	Asia
ERR1620338	54082192	ERS1343400	ERP017091	Healthy	NA	China	Asia
ERR1620339	61997446	ERS1343401	ERP017091	Healthy	NA	China	Asia

ERR1620340	60224946	ERS1343402	ERP017091	Healthy	NA	China	Asia
ERR1620341	58063390	ERS1343403	ERP017091	Healthy	NA	China	Asia
ERR1620342	60155066	ERS1343404	ERP017091	Healthy	NA	China	Asia
ERR1620343	63875982	ERS1343405	ERP017091	Healthy	NA	China	Asia
ERR1620344	54841880	ERS1343406	ERP017091	Healthy	NA	China	Asia
ERR1620345	66805552	ERS1343407	ERP017091	Healthy	NA	China	Asia
ERR1620346	50164918	ERS1343408	ERP017091	Healthy	NA	China	Asia
ERR1620347	57831538	ERS1343409	ERP017091	Healthy	NA	China	Asia
ERR1620348	57028624	ERS1343410	ERP017091	Healthy	NA	China	Asia
ERR1620349	54593268	ERS1343411	ERP017091	Healthy	NA	China	Asia
ERR1620350	59780534	ERS1343412	ERP017091	Healthy	NA	China	Asia
ERR1620351	50858900	ERS1343413	ERP017091	Healthy	NA	China	Asia
ERR1620352	55524646	ERS1343414	ERP017091	Healthy	NA	China	Asia
ERR1620353	59059640	ERS1343415	ERP017091	Healthy	NA	China	Asia
ERR1620354	55450576	ERS1343416	ERP017091	Healthy	NA	China	Asia
ERR1620355	58208070	ERS1343417	ERP017091	Healthy	NA	China	Asia
ERR1620356	56689730	ERS1343418	ERP017091	Healthy	NA	China	Asia
ERR1620357	53488942	ERS1343419	ERP017091	Healthy	NA	China	Asia
ERR1620361	60971404	ERS1343423	ERP017091	Healthy	NA	China	Asia
ERR1620362	71644624	ERS1343424	ERP017091	Healthy	NA	China	Asia
ERR1620363	64300876	ERS1343425	ERP017091	Healthy	NA	China	Asia
ERR1620364	64082020	ERS1343426	ERP017091	Healthy	NA	China	Asia
ERR1620365	50593624	ERS1343427	ERP017091	Healthy	NA	China	Asia
ERR1620367	53844844	ERS1343429	ERP017091	NA	NA	China	Asia
ERR1620368	53392028	ERS1343430	ERP017091	Healthy	NA	China	Asia
ERR1620369	60776632	ERS1343431	ERP017091	Healthy	NA	China	Asia
ERR1620370	60173936	ERS1343432	ERP017091	NA	NA	China	Asia
ERR1620371	52990336	ERS1343433	ERP017091	NA	NA	China	Asia
ERR1620372	58492536	ERS1343434	ERP017091	Healthy	NA	China	Asia
ERR1620373	54191708	ERS1343435	ERP017091	Healthy	NA	China	Asia
ERR1620374	62896290	ERS1343436	ERP017091	Healthy	NA	China	Asia
ERR1620375	56159156	ERS1343437	ERP017091	Healthy	NA	China	Asia
ERR1620376	53882530	ERS1343438	ERP017091	Healthy	NA	China	Asia
ERR1620377	62950410	ERS1343439	ERP017091	Healthy	NA	China	Asia
ERR2013555	74162024	ERS1487525	ERP020710	Diseased	NA	China	Asia
ERR2013556	65876622	ERS1487526	ERP020710	Diseased	NA	China	Asia
ERR2013557	73147654	ERS1487527	ERP020710	Healthy	NA	China	Asia
ERR2013558	70311356	ERS1487528	ERP020710	Healthy	NA	China	Asia
ERR2013559	65638818	ERS1487529	ERP020710	Healthy	NA	China	Asia
ERR2013560	76139652	ERS1487530	ERP020710	Healthy	NA	China	Asia
ERR2013561	60494754	ERS1487531	ERP020710	Healthy	NA	China	Asia
ERR2013562	81496068	ERS1487532	ERP020710	Diseased	NA	China	Asia
ERR2013563	72300276	ERS1487533	ERP020710	Healthy	NA	China	Asia
ERR2013564	65210656	ERS1487534	ERP020710	Healthy	NA	China	Asia
ERR2013565	83929260	ERS1487535	ERP020710	Healthy	NA	China	Asia
ERR2013566	78409710	ERS1487536	ERP020710	Healthy	NA	China	Asia
ERR2013567	73709792	ERS1487537	ERP020710	Healthy	NA	China	Asia
ERR2013568	71494566	ERS1487538	ERP020710	Healthy	NA	China	Asia
ERR2013569	76640318	ERS1487539	ERP020710	Healthy	NA	China	Asia
ERR2013570	70161340	ERS1487540	ERP020710	Healthy	NA	China	Asia
ERR2013577	64373560	ERS1487547	ERP020710	Healthy	NA	China	Asia
ERR2013582	72391598	ERS1487552	ERP020710	Healthy	NA	China	Asia
ERR2013583	55411194	ERS1487553	ERP020710	Healthy	NA	China	Asia
ERR2013584	136830846	ERS1487554	ERP020710	Healthy	NA	China	Asia
ERR2013585	60146294	ERS1487555	ERP020710	Healthy	NA	China	Asia
ERR2013587	74872100	ERS1487557	ERP020710	Healthy	NA	China	Asia
ERR2013593	71971288	ERS1487563	ERP020710	Healthy	NA	China	Asia
ERR2013594	67790558	ERS1487564	ERP020710	Healthy	NA	China	Asia
ERR2013596	66847592	ERS1487566	ERP020710	Healthy	NA	China	Asia
ERR2013599	74550456	ERS1487569	ERP020710	Healthy	NA	China	Asia
ERR2013604	68653330	ERS1487574	ERP020710	Diseased	NA	China	Asia
ERR2013605	67772568	ERS1487575	ERP020710	Healthy	NA	China	Asia
ERR2013607	73893004	ERS1487577	ERP020710	Healthy	NA	China	Asia
ERR2013609	77604004	ERS1487579	ERP020710	Healthy	NA	China	Asia
ERR2013610	61792236	ERS1487580	ERP020710	Healthy	NA	China	Asia
ERR2013611	77333508	ERS1487581	ERP020710	Healthy	NA	China	Asia
ERR2013614	55625268	ERS1487584	ERP020710	Healthy	NA	China	Asia
ERR2013616	52805300	ERS1487586	ERP020710	Healthy	NA	China	Asia
ERR2013617	55424562	ERS1487587	ERP020710	Healthy	NA	China	Asia
ERR2013618	125703958	ERS1487588	ERP020710	Diseased	NA	China	Asia
ERR2013619	60316794	ERS1487589	ERP020710	Diseased	NA	China	Asia
ERR2013622	81498014	ERS1487592	ERP020710	Healthy	NA	China	Asia

ERR2013623	53615460	ERS1487593	ERP020710	Healthy	NA	China	Asia
ERR2013624	90827984	ERS1487594	ERP020710	Healthy	NA	China	Asia
ERR2013628	51753894	ERS1487598	ERP020710	Healthy	NA	China	Asia
ERR2013636	61773236	ERS1487606	ERP020710	Healthy	NA	China	Asia
ERR2013637	60147830	ERS1487607	ERP020710	Healthy	NA	China	Asia
ERR2013639	53678738	ERS1487609	ERP020710	Healthy	NA	China	Asia
ERR2013644	57680918	ERS1487614	ERP020710	Healthy	NA	China	Asia
ERR2013647	68306306	ERS1487617	ERP020710	Healthy	NA	China	Asia
ERR2013649	54902848	ERS1487619	ERP020710	Diseased	NA	China	Asia
ERR2013651	60034738	ERS1487621	ERP020710	Healthy	NA	China	Asia
ERR2013652	54300674	ERS1487622	ERP020710	Healthy	NA	China	Asia
ERR2013653	53714630	ERS1487623	ERP020710	Healthy	NA	China	Asia
ERR2013654	53822476	ERS1487624	ERP020710	Diseased	NA	China	Asia
ERR1953518	77164794	ERS1295930	ERP022699	Healthy	Urban	United Kingdom	Europe
ERR1953519	73616924	ERS1295851	ERP022699	Healthy	Urban	United Kingdom	Europe
ERR1953520	89044282	ERS1282030	ERP022699	Healthy	Urban	United Kingdom	Europe
ERR1953521	90183212	ERS1282031	ERP022699	Healthy	Urban	United Kingdom	Europe
ERR1953522	90058092	ERS1282032	ERP022699	Healthy	Urban	United Kingdom	Europe
ERR1953523	99692758	ERS1282033	ERP022699	Healthy	Urban	United Kingdom	Europe
SRR1778451	58533772	SRS333665	SRP008047	Diseased	NA	NA	Asia
SRR1778452	65516874	SRS333666	SRP008047	Diseased	NA	NA	Asia
SRR1778453	56129076	SRS333667	SRP008047	Diseased	NA	NA	Asia
SRR1778454	79916450	SRS333668	SRP008047	Diseased	NA	NA	Asia
SRR1778455	99784820	SRS333669	SRP008047	Diseased	NA	NA	Asia
SRR1778456	77694220	SRS333670	SRP008047	Diseased	NA	NA	Asia
SRR341632	61005902	SRS259485	SRP008047	Healthy	NA	China	Asia
SRR341633	61163528	SRS259486	SRP008047	Healthy	NA	China	Asia
SRR413559	52442888	SRS294815	SRP008047	NA	NA	NA	NA
SRR413566	50882972	SRS294822	SRP008047	NA	NA	NA	NA
SRR413568	63521966	SRS294824	SRP008047	NA	NA	NA	NA
SRR413571	50627868	SRS294827	SRP008047	NA	NA	NA	NA
SRR413573	61965310	SRS294829	SRP008047	NA	NA	NA	NA
SRR413605	72548932	SRS294861	SRP008047	NA	NA	NA	NA
SRR413607	53911072	SRS294863	SRP008047	NA	NA	NA	NA
SRR413609	52649798	SRS294865	SRP008047	NA	NA	NA	NA
SRR413619	51323938	SRS294875	SRP008047	NA	NA	NA	NA
SRR413620	51481706	SRS294876	SRP008047	NA	NA	NA	NA
SRR413634	51414960	SRS294890	SRP008047	Healthy	NA	NA	Asia
SRR413635	50281118	SRS294891	SRP008047	NA	NA	NA	NA
SRR413660	59795072	SRS294916	SRP008047	NA	NA	NA	NA
SRR413663	58617474	SRS294919	SRP008047	NA	NA	NA	NA
SRR413664	56580850	SRS294920	SRP008047	NA	NA	NA	NA
SRR413674	53980064	SRS294930	SRP008047	Diseased	NA	China	Asia
SRR413675	59355112	SRS294931	SRP008047	Diseased	NA	China	Asia
SRR413677	64697348	SRS294933	SRP008047	Diseased	NA	China	Asia
SRR413678	69615406	SRS294934	SRP008047	NA	NA	NA	NA
SRR413679	55243200	SRS294935	SRP008047	NA	NA	NA	NA
SRR413682	57536000	SRS294938	SRP008047	NA	NA	NA	NA
SRR413683	51405986	SRS294939	SRP008047	NA	NA	NA	NA
SRR413686	55630074	SRS294942	SRP008047	Diseased	NA	China	Asia
SRR413688	57833192	SRS294944	SRP008047	NA	NA	NA	NA
SRR413689	59600380	SRS294945	SRP008047	NA	NA	NA	NA
SRR413690	53958068	SRS294946	SRP008047	NA	NA	NA	NA
SRR413692	51845286	SRS294948	SRP008047	NA	NA	NA	NA
SRR413693	57356636	SRS294949	SRP008047	NA	NA	NA	NA
SRR413694	56442720	SRS294950	SRP008047	Diseased	NA	China	Asia
SRR413695	58503804	SRS294951	SRP008047	NA	NA	NA	NA
SRR413698	58333542	SRS294954	SRP008047	NA	NA	NA	NA
SRR413700	57831770	SRS294956	SRP008047	Diseased	NA	China	Asia
SRR413703	51021308	SRS294959	SRP008047	NA	NA	NA	NA
SRR413705	53692560	SRS294961	SRP008047	NA	NA	NA	NA
SRR413706	58242046	SRS294962	SRP008047	NA	NA	NA	NA
SRR413708	67734538	SRS294964	SRP008047	Diseased	NA	China	Asia
SRR413710	58275484	SRS294966	SRP008047	NA	NA	NA	NA
SRR413714	54470488	SRS294970	SRP008047	Diseased	NA	China	Asia
SRR413715	75983422	SRS294971	SRP008047	NA	NA	NA	NA
SRR413716	72503174	SRS294972	SRP008047	NA	NA	NA	NA
SRR413721	55908886	SRS294977	SRP008047	Diseased	NA	China	Asia
SRR413722	62538360	SRS294978	SRP008047	NA	NA	NA	NA
SRR413726	51430690	SRS294982	SRP008047	Diseased	NA	NA	Asia
SRR413727	66066136	SRS294983	SRP008047	Diseased	NA	NA	Asia
SRR413728	59652908	SRS294984	SRP008047	Diseased	NA	China	Asia

SRR413733	68858570	SRS294989	SRP008047	Diseased	NA	China	Asia
SRR413735	53652434	SRS294991	SRP008047	Diseased	NA	China	Asia
SRR413736	60707470	SRS294992	SRP008047	Diseased	NA	China	Asia
SRR413737	57149066	SRS294993	SRP008047	Diseased	NA	China	Asia
SRR413739	54093762	SRS294995	SRP008047	NA	NA	China	Asia
SRR413740	52213242	SRS294996	SRP008047	NA	NA	China	Asia
SRR413750	50447092	SRS295006	SRP008047	Diseased	NA	China	Asia
SRR413753	61312836	SRS295009	SRP008047	NA	NA	NA	NA
SRR413754	76797816	SRS295010	SRP008047	Diseased	NA	China	Asia
SRR413757	66440864	SRS295013	SRP008047	Diseased	NA	China	Asia
SRR413760	50226912	SRS295016	SRP008047	Diseased	NA	China	Asia
SRR413761	66835248	SRS295017	SRP008047	Diseased	NA	China	Asia
SRR413762	62012558	SRS295018	SRP008047	Diseased	NA	China	Asia
SRR413764	51594418	SRS295020	SRP008047	Diseased	NA	China	Asia
SRR413765	52225992	SRS295021	SRP008047	Diseased	NA	China	Asia
SRR413766	71108892	SRS295022	SRP008047	Diseased	NA	China	Asia
SRR413768	57338014	SRS295024	SRP008047	Diseased	NA	China	Asia
SRR413769	72044412	SRS295025	SRP008047	Diseased	NA	China	Asia
SRR413770	54549592	SRS295026	SRP008047	Diseased	NA	China	Asia
SRR413773	62432024	SRS295029	SRP008047	Diseased	NA	China	Asia
SRR453563	56536658	SRS307175	SRP012035	Diseased	NA	NA	NA
SRR453564	59294032	SRS308056	SRP012035	Diseased	NA	NA	NA
SRR453565	59322120	SRS307177	SRP012035	Diseased	NA	NA	NA
SRR2223198	51662092	SRS477428	SRP029441	Healthy	NA	Fiji	Oceania
SRR2223207	51844574	SRS477428	SRP029441	Healthy	NA	Fiji	Oceania
SRR2223229	51959612	SRS476326	SRP029441	Healthy	NA	Fiji	Oceania
SRR2223242	52839686	SRS476326	SRP029441	Healthy	NA	Fiji	Oceania
SRR2223418	52611832	SRS476326	SRP029441	Healthy	NA	Fiji	Oceania
SRR2223495	51323324	SRS477428	SRP029441	Healthy	NA	Fiji	Oceania
SRR2223515	52953636	SRS476326	SRP029441	Healthy	NA	Fiji	Oceania
SRR2226375	56446010	SRS475925	SRP029441	Healthy	NA	Fiji	Oceania
SRR2227557	50335186	SRS475568	SRP029441	Healthy	NA	Fiji	Oceania
SRR2227597	56633710	SRS475925	SRP029441	Healthy	NA	Fiji	Oceania
SRR2227601	56406462	SRS475925	SRP029441	Healthy	NA	Fiji	Oceania
SRR2227650	50119564	SRS475568	SRP029441	Healthy	NA	Fiji	Oceania
SRR2227861	51692062	SRS477428	SRP029441	Healthy	NA	Fiji	Oceania
SRR2228651	56358178	SRS475925	SRP029441	Healthy	NA	Fiji	Oceania
SRR2228802	50049426	SRS475568	SRP029441	Healthy	NA	Fiji	Oceania
SRR2244733	53325066	SRS476342	SRP029441	Healthy	NA	Fiji	Oceania
SRR2245052	53802368	SRS476342	SRP029441	Healthy	NA	Fiji	Oceania
SRR2250411	53773568	SRS476342	SRP029441	Healthy	NA	Fiji	Oceania
SRR2250454	53798064	SRS476342	SRP029441	Healthy	NA	Fiji	Oceania
SRR1039532	197716264	SRS508588	SRP033353	Diseased	Urban	United States	North America
SRR1039533	141148978	SRS508590	SRP033353	Diseased	Urban	United States	North America
SRR1761677	51851920	SRS820585	SRP052307	Healthy	Urban	United States	North America
SRR1761678	53588754	SRS820586	SRP052307	Healthy	Urban	United States	North America
SRR1761690	56587752	SRS820598	SRP052307	Healthy	Urban	United States	North America
SRR1761697	53139708	SRS820605	SRP052307	Healthy	Urban	United States	North America
SRR1761698	70798988	SRS820606	SRP052307	Healthy	NA	Peru	South America
SRR1761699	90047818	SRS820607	SRP052307	Healthy	NA	Peru	South America
SRR1761702	64274686	SRS820610	SRP052307	Healthy	NA	Peru	South America
SRR1761703	62225764	SRS820614	SRP052307	Healthy	NA	Peru	South America
SRR1761704	58107124	SRS820611	SRP052307	Healthy	NA	Peru	South America
SRR1761705	58453308	SRS820612	SRP052307	Healthy	NA	Peru	South America
SRR1761706	64246800	SRS820613	SRP052307	Healthy	NA	Peru	South America
SRR1761707	61151082	SRS820616	SRP052307	Healthy	NA	Peru	South America
SRR1761708	53826426	SRS820615	SRP052307	Healthy	NA	Peru	South America
SRR1761709	59296106	SRS820617	SRP052307	Healthy	NA	Peru	South America
SRR1761710	54795168	SRS820618	SRP052307	Healthy	NA	Peru	South America
SRR1761711	59507584	SRS820619	SRP052307	Healthy	NA	Peru	South America
SRR1761712	68630212	SRS820620	SRP052307	Healthy	NA	Peru	South America
SRR1761713	56013838	SRS820621	SRP052307	Healthy	NA	Peru	South America
SRR1761714	53025866	SRS820622	SRP052307	Healthy	NA	Peru	South America
SRR1761715	53759636	SRS820623	SRP052307	Healthy	NA	Peru	South America
SRR1761716	66278546	SRS820624	SRP052307	Healthy	NA	Peru	South America
SRR1761717	63999480	SRS820625	SRP052307	Healthy	NA	Peru	South America
SRR1761718	59981100	SRS820626	SRP052307	Healthy	NA	Peru	South America
SRR1761719	55679472	SRS820627	SRP052307	Healthy	NA	Peru	South America
SRR1761720	58169924	SRS820628	SRP052307	Healthy	NA	Peru	South America
SRR1761721	58106422	SRS820629	SRP052307	Healthy	NA	Peru	South America
SRR1779472	90056400	SRS824302	SRP052424	Healthy	NA	Singapore	Asia
SRR1783777	101649340	SRS824303	SRP052424	Healthy	NA	Singapore	Asia

SRR1789035	77913270	SRS827335	SRP052424	Diseased	NA	Singapore	Asia
SRR1793377	67215026	SRS828776	SRP052424	Diseased	NA	Singapore	Asia
SRR1793416	82821416	SRS820971	SRP052424	Healthy	NA	Singapore	Asia
SRR1799892	96515272	SRS839650	SRP052424	Healthy	NA	Singapore	Asia
SRR1819806	73735880	SRS845106	SRP052424	Healthy	NA	Singapore	Asia
SRR1821190	91813672	SRS857060	SRP052424	Diseased	NA	Singapore	Asia
SRR1822318	79717456	SRS857061	SRP052424	Diseased	NA	Singapore	Asia
SRR1825362	95294552	SRS859975	SRP052424	Diseased	NA	Singapore	Asia
SRR1825367	95267886	SRS862521	SRP052424	Diseased	NA	Singapore	Asia
SRR1929408	63686272	SRS882242	SRP056480	Healthy	NA	United Republic of Tanzania	Africa
SRR1930121	71290650	SRS883019	SRP056480	Healthy	NA	United Republic of Tanzania	Africa
SRR1930123	77841428	SRS883021	SRP056480	Healthy	NA	United Republic of Tanzania	Africa
SRR1930141	64739338	SRS883029	SRP056480	Healthy	NA	United Republic of Tanzania	Africa
SRR2047620	233353868	SRS935438	SRP058320	NA	NA	United States	North America
SRR2047845	181672134	SRS935447	SRP058320	Diseased	Urban	United States	North America
SRR2048044	70333202	SRS935445	SRP058320	Diseased	Urban	United States	North America
SRR2048045	69876892	SRS935445	SRP058320	Diseased	Urban	United States	North America
SRR5050591	63978278	SRS1816457	SRP058320	NA	NA	United States	North America
SRR5050592	64809984	SRS1816457	SRP058320	NA	NA	United States	North America
SRR2164314	273294044	SRS1035613	SRP060278	NA	NA	NA	NA
SRR2155338	54135726	SRS1028198	SRP062282	Healthy	Urban	Germany	Europe
SRR2155482	114595596	SRS1028549	SRP062282	Diseased	Urban	Germany	Europe
SRR2673277	82571062	SRS1117397	SRP064913	Healthy	Urban	United States	North America
SRR2673278	184614246	SRS1117397	SRP064913	Healthy	Urban	United States	North America
SRR2673315	75177634	SRS1117397	SRP064913	Healthy	Urban	United States	North America
SRR2674233	80014180	SRS1117397	SRP064913	Healthy	Urban	United States	North America
SRR2674234	190249176	SRS1117397	SRP064913	Healthy	Urban	United States	North America
SRR2674235	147902952	SRS1117397	SRP064913	Healthy	Urban	United States	North America
SRR2725846	188760530	SRS1117522	SRP064913	Healthy	Urban	United States	North America
SRR2725847	168095640	SRS1117522	SRP064913	Healthy	Urban	United States	North America
SRR2725850	74067160	SRS1117522	SRP064913	Healthy	Urban	United States	North America
SRR2725928	75707006	SRS1117522	SRP064913	Healthy	Urban	United States	North America
SRR2726028	124403594	SRS1117522	SRP064913	Healthy	Urban	United States	North America
SRR2726047	102595918	SRS1117522	SRP064913	Healthy	Urban	United States	North America
SRR2726135	153804104	SRS1117528	SRP064913	Healthy	Urban	United States	North America
SRR2726136	85420060	SRS1117528	SRP064913	Healthy	Urban	United States	North America
SRR2726240	65489780	SRS1117528	SRP064913	Healthy	Urban	United States	North America
SRR2726241	70108028	SRS1117528	SRP064913	Healthy	Urban	United States	North America
SRR2726242	101317812	SRS1117528	SRP064913	Healthy	Urban	United States	North America
SRR2726243	113355732	SRS1117528	SRP064913	Healthy	Urban	United States	North America
SRR2726244	64484820	SRS1117537	SRP064913	Healthy	Urban	United States	North America
SRR2726248	69332162	SRS1117537	SRP064913	Healthy	Urban	United States	North America
SRR2726600	75873740	SRS1117537	SRP064913	Healthy	Urban	United States	North America
SRR2726601	76886248	SRS1117537	SRP064913	Healthy	Urban	United States	North America
SRR2726602	121552398	SRS1117537	SRP064913	Healthy	Urban	United States	North America
SRR2726603	90109816	SRS1117537	SRP064913	Healthy	Urban	United States	North America
SRR2846706	498756462	SRS1133497	SRP065270	Healthy	NA	United States	North America
SRR2857332	126204562	SRS1135680	SRP065270	Healthy	NA	United States	North America
SRR2857686	111336336	SRS1135668	SRP065270	Healthy	Urban	United States	North America
SRR2857885	139903424	SRS1135683	SRP065270	Healthy	Urban	United States	North America
SRR2857886	137610596	SRS1135709	SRP065270	Healthy	Urban	United States	North America
SRR2857969	135898360	SRS1135723	SRP065270	Healthy	NA	United States	North America
SRR2857970	124444716	SRS1135724	SRP065270	Healthy	Urban	United States	North America
SRR2858047	119429514	SRS1135725	SRP065270	Healthy	NA	United States	North America
SRR2858128	98589216	SRS1135768	SRP065270	Healthy	Urban	United States	North America
SRR2912777	53575258	SRS1158738	SRP066053	Diseased	NA	United States	North America
SRR2912779	54602024	SRS1158736	SRP066053	Diseased	NA	United States	North America
SRR2912787	54875512	SRS1158752	SRP066053	Diseased	NA	United States	North America
SRR2912789	54068546	SRS1158749	SRP066053	Diseased	NA	United States	North America
SRR2912799	60284686	SRS1158759	SRP066053	Healthy	NA	United States	North America
SRR2912800	53718300	SRS1158758	SRP066053	Diseased	NA	United States	North America
SRR2912801	61659748	SRS1158757	SRP066053	Diseased	NA	United States	North America
SRR2912802	50874934	SRS1158756	SRP066053	Diseased	NA	United States	North America
SRR2912803	54417132	SRS1158755	SRP066053	Diseased	NA	United States	North America
SRR2912804	52690478	SRS1158754	SRP066053	Diseased	NA	United States	North America
SRR3108049	53100274	SRS1253248	SRP068612	NA	Urban	Ireland	Europe
SRR3108056	53184294	SRS1253247	SRP068612	NA	Urban	Ireland	Europe
SRR3108064	53187018	SRS1253251	SRP068612	NA	Urban	Ireland	Europe
SRR3108072	53306868	SRS1253250	SRP068612	NA	Urban	Ireland	Europe
SRR3108079	53274964	SRS1253252	SRP068612	NA	Urban	Ireland	Europe
SRR3110877	53032280	SRS1253249	SRP068612	NA	Urban	Ireland	Europe
SRR3160438	103433114	SRS1283432	SRP069867	Healthy	Urban	United States	North America

SRR3160439	99283572	SRS1283433	SRP069867	Healthy	Urban	United States	North America
SRR3160443	55880152	SRS1283435	SRP069867	Diseased	Urban	United States	North America
SRR3160444	55467566	SRS1283435	SRP069867	Diseased	Urban	United States	North America
SRR3160452	64519826	SRS1283440	SRP069867	Diseased	Urban	United States	North America
SRR3160453	64046756	SRS1283440	SRP069867	Diseased	Urban	United States	North America
SRR3160454	83194242	SRS1283476	SRP069867	Diseased	Urban	United States	North America
SRR3160455	53800066	SRS1283477	SRP069867	Diseased	Urban	United States	North America
SRR3160456	53451452	SRS1283477	SRP069867	Diseased	Urban	United States	North America
SRR3160459	91851220	SRS1283597	SRP069867	Diseased	Urban	United States	North America
SRR3160460	99099426	SRS1283617	SRP069867	Diseased	Urban	United States	North America
SRR3195484	169229076	SRS1315473	SRP070971	NA	NA	United States	North America
SRR3340629	64108898	SRS1378921	SRP072916	NA	NA	Germany	Europe
SRR3340631	55717278	SRS1378922	SRP072916	NA	NA	Germany	Europe
SRR3466404	491146884	SRS1417035	SRP074153	Diseased	Urban	United States	North America
SRR3498907	282529954	SRS1433906	SRP074153	Diseased	Urban	United States	North America
SRR3498909	154320678	SRS1433907	SRP074153	Diseased	Urban	United States	North America
SRR3506419	448318652	SRS1433910	SRP074153	Diseased	Urban	United States	North America
SRR3506420	767799042	SRS1437768	SRP074153	NA	NA	United States	North America
SRR3546776	329237490	SRS1446905	SRP074153	Diseased	Urban	United States	North America
SRR3546778	99566746	SRS1446907	SRP074153	Diseased	Urban	United States	North America
SRR3546779	616411722	SRS1446908	SRP074153	Diseased	Urban	United States	North America
SRR3546780	700979492	SRS1446909	SRP074153	Diseased	Urban	United States	North America
SRR3546781	447285180	SRS1446910	SRP074153	Diseased	Urban	United States	North America
SRR3546782	418189516	SRS1446911	SRP074153	Diseased	Urban	United States	North America
SRR6257422	84654038	SRS1446908	SRP074153	Diseased	Urban	United States	North America
SRR6257423	50776458	SRS1446908	SRP074153	Diseased	Urban	United States	North America
SRR6257426	69652444	SRS1446908	SRP074153	Diseased	Urban	United States	North America
SRR6257455	72048496	SRS1446908	SRP074153	Diseased	Urban	United States	North America
SRR6257457	91278422	SRS1446908	SRP074153	Diseased	Urban	United States	North America
SRR6257458	63479076	SRS1446908	SRP074153	Diseased	Urban	United States	North America
SRR6257489	82538854	SRS1437768	SRP074153	Diseased	Urban	United States	North America
SRR6257494	82047114	SRS1437768	SRP074153	Diseased	Urban	United States	North America
SRR6257510	71299036	SRS1437768	SRP074153	Diseased	Urban	United States	North America
SRR6257511	86804068	SRS1437768	SRP074153	Diseased	Urban	United States	North America
SRR6257515	68407838	SRS1437768	SRP074153	Diseased	Urban	United States	North America
SRR3496379	55526980	SRS1432719	SRP074801	Healthy	Urban	United Kingdom	Europe
SRR3582131	53678536	SRS1465228	SRP075633	Diseased	Urban	United States	North America
SRR3582136	53640580	SRS1465233	SRP075633	Diseased	Urban	United States	North America
SRR3582144	51558134	SRS1465242	SRP075633	Diseased	Urban	United States	North America
SRR3582148	54143966	SRS1465245	SRP075633	Diseased	Urban	United States	North America
SRR3582150	71640096	SRS1465247	SRP075633	Diseased	Urban	United States	North America
SRR3582151	50912530	SRS1465249	SRP075633	Diseased	Urban	United States	North America
SRR3582152	71712498	SRS1465248	SRP075633	Diseased	Urban	United States	North America
SRR3582153	74132932	SRS1465250	SRP075633	Diseased	Urban	United States	North America
SRR3582155	64886912	SRS1465252	SRP075633	Diseased	Urban	United States	North America
SRR3582157	59760956	SRS1465254	SRP075633	Diseased	Urban	United States	North America
SRR3582158	59219494	SRS1465255	SRP075633	Healthy	Urban	United States	North America
SRR3582159	69667608	SRS1465256	SRP075633	Diseased	Urban	United States	North America
SRR3582160	72450648	SRS1465257	SRP075633	Diseased	Urban	United States	North America
SRR3582162	60658592	SRS1465259	SRP075633	Diseased	Urban	United States	North America
SRR3582163	62317186	SRS1465260	SRP075633	Diseased	Urban	United States	North America
SRR3582164	69641672	SRS1465261	SRP075633	Diseased	Urban	United States	North America
SRR3582165	62980280	SRS1465263	SRP075633	Diseased	Urban	United States	North America
SRR3582168	65914032	SRS1465265	SRP075633	Diseased	Urban	United States	North America
SRR3582169	52601574	SRS1465267	SRP075633	Diseased	Urban	United States	North America
SRR3582174	51488948	SRS1465270	SRP075633	Diseased	Urban	United States	North America
SRR3582176	58511776	SRS1465273	SRP075633	Diseased	Urban	United States	North America
SRR3582177	96137588	SRS1465274	SRP075633	Diseased	Urban	United States	North America
SRR3582179	94990870	SRS1465276	SRP075633	NA	NA	United States	North America
SRR3582181	74110592	SRS1465278	SRP075633	Diseased	Urban	United States	North America
SRR3582182	53355220	SRS1465279	SRP075633	Diseased	Urban	United States	North America
SRR3737021	73164788	SRS1490018	SRP076119	Healthy	Urban	United States	North America
SRR3917562	68487484	SRS1563115	SRP076119	Healthy	Urban	United States	North America
SRR3917627	57275068	SRS1563124	SRP076119	Healthy	Urban	United States	North America
SRR3917687	59571700	SRS1563130	SRP076119	Healthy	Urban	United States	North America
SRR3992955	78376762	SRS1596768	SRP080787	Healthy	NA	Mongolia	Asia
SRR3992958	54269420	SRS1596771	SRP080787	Healthy	NA	China	Asia
SRR3992959	64648644	SRS1596772	SRP080787	Healthy	NA	Mongolia	Asia
SRR3992961	82546276	SRS1596774	SRP080787	Healthy	NA	China	Asia
SRR3992962	65661186	SRS1596775	SRP080787	Healthy	NA	China	Asia
SRR3992965	88086108	SRS1596778	SRP080787	Healthy	NA	Mongolia	Asia
SRR3992967	62844638	SRS1596780	SRP080787	Healthy	NA	China	Asia

SRR3992969	107831094	SRS1596782	SRP080787	Healthy	NA	Mongolia	Asia
SRR3992971	62024590	SRS1596784	SRP080787	Healthy	NA	Mongolia	Asia
SRR3992973	130254616	SRS1596785	SRP080787	Healthy	NA	Mongolia	Asia
SRR3992978	63939470	SRS1596791	SRP080787	Healthy	NA	Mongolia	Asia
SRR3992980	50954622	SRS1596793	SRP080787	Healthy	NA	Mongolia	Asia
SRR3992981	58568686	SRS1596794	SRP080787	Healthy	NA	Mongolia	Asia
SRR3992984	50284672	SRS1596796	SRP080787	Healthy	NA	Mongolia	Asia
SRR3992985	110446032	SRS1596798	SRP080787	Healthy	NA	Mongolia	Asia
SRR3992987	181557884	SRS1596800	SRP080787	Healthy	NA	Mongolia	Asia
SRR3992990	67393674	SRS1596803	SRP080787	Healthy	NA	Mongolia	Asia
SRR3992991	193607438	SRS1596804	SRP080787	Healthy	NA	Mongolia	Asia
SRR3992993	72036280	SRS1596806	SRP080787	Healthy	NA	Mongolia	Asia
SRR3992995	77807568	SRS1596808	SRP080787	Healthy	NA	Mongolia	Asia
SRR3992997	70848376	SRS1596809	SRP080787	Healthy	NA	Mongolia	Asia
SRR3992998	68869128	SRS1596811	SRP080787	Healthy	NA	Mongolia	Asia
SRR3992999	85952440	SRS1596812	SRP080787	Healthy	NA	Mongolia	Asia
SRR3993000	56078692	SRS1596813	SRP080787	Healthy	NA	Mongolia	Asia
SRR3993001	57493562	SRS1596814	SRP080787	Healthy	NA	Mongolia	Asia
SRR3993002	60532296	SRS1596815	SRP080787	Healthy	NA	Mongolia	Asia
SRR3993003	52714170	SRS1596816	SRP080787	Healthy	NA	Mongolia	Asia
SRR3993012	79729538	SRS1596824	SRP080787	Healthy	NA	Mongolia	Asia
SRR3993013	72365638	SRS1596826	SRP080787	Healthy	NA	Mongolia	Asia
SRR3993014	65522780	SRS1596827	SRP080787	Healthy	NA	Mongolia	Asia
SRR3993018	63578668	SRS1596831	SRP080787	Healthy	NA	Mongolia	Asia
SRR3993023	60882172	SRS1596836	SRP080787	Healthy	NA	Mongolia	Asia
SRR3993030	75342254	SRS1596843	SRP080787	Healthy	NA	Mongolia	Asia
SRR3993040	72434026	SRS1596853	SRP080787	Healthy	NA	Mongolia	Asia
SRR3993042	80753778	SRS1596855	SRP080787	Healthy	NA	Mongolia	Asia
SRR3993046	58766464	SRS1596859	SRP080787	Healthy	NA	China	Asia
SRR3993047	50485574	SRS1596860	SRP080787	Healthy	NA	China	Asia
SRR3993060	73467432	SRS1596873	SRP080787	Healthy	NA	China	Asia
SRR4033070	85373492	SRS1618830	SRP082182	NA	NA	United States	North America
SRR4033072	55222024	SRS1618832	SRP082182	Healthy	Urban	United States	North America
SRR4033074	72414262	SRS1618834	SRP082182	Diseased	Urban	United States	North America
SRR4033075	51617272	SRS1618835	SRP082182	Diseased	Urban	United States	North America
SRR4052025	50370956	SRS1634638	SRP082656	Healthy	NA	Italy	Europe
SRR4305187	58008948	SRS1719244	SRP090628	Healthy	NA	Russia	Europe
SRR4305222	54192890	SRS1719278	SRP090628	Healthy	NA	Russia	Europe
SRR4305267	52185144	SRS1719322	SRP090628	Healthy	NA	Russia	Europe
SRR4305405	71175562	SRS1719457	SRP090628	Healthy	NA	Russia	Europe
SRR4305482	53623114	SRS1719535	SRP090628	Healthy	NA	Russia	Europe
SRR4408074	50206866	SRS1735509	SRP090628	Healthy	NA	Estonia	Europe
SRR4408150	57342704	SRS1735579	SRP090628	Healthy	NA	Russia	Europe
SRR4408152	51893108	SRS1735582	SRP090628	Healthy	NA	Finland	Europe
SRR4408211	62482038	SRS1735640	SRP090628	Healthy	NA	Russia	Europe
SRR4420318	58450870	SRS1743808	SRP091494	NA	NA	NA	NA
SRR4423578	79211006	SRS1746270	SRP091570	Diseased	Urban	United States	North America
SRR4423579	87107690	SRS1746271	SRP091570	NA	Urban	United States	North America
SRR4423581	83413608	SRS1746272	SRP091570	Diseased	Urban	United States	North America
SRR4423616	86363926	SRS1746273	SRP091570	Diseased	Urban	United States	North America
SRR4423631	95573752	SRS1746275	SRP091570	Diseased	Urban	United States	North America
SRR4423633	91029528	SRS1746277	SRP091570	NA	Urban	United States	North America
SRR4423642	92041320	SRS1746278	SRP091570	Diseased	Urban	United States	North America
SRR4423656	90705538	SRS1746274	SRP091570	NA	Urban	United States	North America
SRR4423662	86320814	SRS1746276	SRP091570	Diseased	Urban	United States	North America
SRR4423675	87665642	SRS1746279	SRP091570	NA	Urban	United States	North America
SRR4423685	74655838	SRS1746281	SRP091570	NA	Urban	United States	North America
SRR4423697	92011450	SRS1746269	SRP091570	Diseased	Urban	United States	North America
SRR4423704	85118934	SRS1746280	SRP091570	NA	Urban	United States	North America
SRR4435697	78827256	SRS1754483	SRP091570	NA	Urban	United States	North America
SRR4435698	82899662	SRS1754484	SRP091570	NA	Urban	United States	North America
SRR4435717	89803844	SRS1754481	SRP091570	NA	Urban	United States	North America
SRR4435731	80737794	SRS1754490	SRP091570	NA	Urban	United States	North America
SRR4435733	68108614	SRS1754489	SRP091570	NA	Urban	United States	North America
SRR4435736	70633110	SRS1754488	SRP091570	NA	Urban	United States	North America
SRR4435750	85043440	SRS1754487	SRP091570	NA	Urban	United States	North America
SRR4435761	84397696	SRS1754485	SRP091570	NA	Urban	United States	North America
SRR4435767	89821426	SRS1754491	SRP091570	NA	Urban	United States	North America
SRR4435785	113790098	SRS1754482	SRP091570	NA	Urban	United States	North America
SRR4435795	87105664	SRS1754480	SRP091570	NA	Urban	United States	North America
SRR4435801	82827774	SRS1754494	SRP091570	NA	Urban	United States	North America
SRR4435814	77156216	SRS1754495	SRP091570	NA	Urban	United States	North America

SRR4444749	72976922	SRS1756239	SRP091570	NA	Urban	United States	North America
SRR4444755	81329028	SRS1756242	SRP091570	Diseased	Urban	United States	North America
SRR4444763	84051000	SRS1756243	SRP091570	NA	Urban	United States	North America
SRR4444766	72648498	SRS1756244	SRP091570	NA	Urban	United States	North America
SRR4444778	75109942	SRS1756246	SRP091570	NA	Urban	United States	North America
SRR4444783	72236398	SRS1756245	SRP091570	NA	Urban	United States	North America
SRR4444801	106438510	SRS1756248	SRP091570	NA	Urban	United States	North America
SRR4444805	75024872	SRS1756249	SRP091570	NA	Urban	United States	North America
SRR4444814	72548668	SRS1756250	SRP091570	NA	Urban	United States	North America
SRR4444820	75180634	SRS1756247	SRP091570	NA	Urban	United States	North America
SRR4444825	81208090	SRS1756240	SRP091570	NA	Urban	United States	North America
SRR4444836	69488294	SRS1756241	SRP091570	Diseased	Urban	United States	North America
SRR4444845	78750284	SRS1756252	SRP091570	Diseased	Urban	United States	North America
SRR4444847	81920994	SRS1754486	SRP091570	Diseased	Urban	United States	North America
SRR4444859	83856686	SRS1756251	SRP091570	Diseased	Urban	United States	North America
SRR4444875	70285332	SRS1756253	SRP091570	NA	Urban	United States	North America
SRR4451535	89855622	SRS1760591	SRP091570	Diseased	Urban	United States	North America
SRR4451543	82523038	SRS1760592	SRP091570	NA	Urban	United States	North America
SRR4451547	84592304	SRS1760593	SRP091570	NA	Urban	United States	North America
SRR4451556	76888268	SRS1760594	SRP091570	Diseased	Urban	United States	North America
SRR4451562	74939132	SRS1760595	SRP091570	NA	Urban	United States	North America
SRR4451579	77842000	SRS1760598	SRP091570	NA	Urban	United States	North America
SRR4451581	87018582	SRS1760599	SRP091570	NA	Urban	United States	North America
SRR4451583	112381110	SRS1760600	SRP091570	NA	Urban	United States	North America
SRR4451585	74616502	SRS1760597	SRP091570	NA	Urban	United States	North America
SRR4451605	76985292	SRS1760601	SRP091570	Diseased	Urban	United States	North America
SRR4451615	75331068	SRS1760602	SRP091570	NA	Urban	United States	North America
SRR4451622	86398010	SRS1760604	SRP091570	Diseased	Urban	United States	North America
SRR4451628	74169422	SRS1760590	SRP091570	Diseased	Urban	United States	North America
SRR4451631	130442806	SRS1760590	SRP091570	Diseased	Urban	United States	North America
SRR4451639	80053562	SRS1760603	SRP091570	NA	Urban	United States	North America
SRR4451646	71549170	SRS1760606	SRP091570	Diseased	Urban	United States	North America
SRR4451660	87927718	SRS1760596	SRP091570	Diseased	Urban	United States	North America
SRR4481704	71381338	SRS1767349	SRP091570	Diseased	Urban	United States	North America
SRR4481706	80133922	SRS1767350	SRP091570	Diseased	Urban	United States	North America
SRR4481717	73284278	SRS1767353	SRP091570	NA	Urban	United States	North America
SRR4481725	74800662	SRS1767354	SRP091570	NA	Urban	United States	North America
SRR4481730	160117904	SRS1767351	SRP091570	Diseased	Urban	United States	North America
SRR4481738	182432370	SRS1767355	SRP091570	NA	Urban	United States	North America
SRR4481744	81887680	SRS1767356	SRP091570	NA	Urban	United States	North America
SRR4481747	92652428	SRS1767357	SRP091570	NA	Urban	United States	North America
SRR4481761	70710678	SRS1767348	SRP091570	Diseased	Urban	United States	North America
SRR4481765	80703176	SRS1767359	SRP091570	NA	Urban	United States	North America
SRR4481767	85627236	SRS1767360	SRP091570	NA	Urban	United States	North America
SRR4481769	71120108	SRS1767361	SRP091570	NA	Urban	United States	North America
SRR4481774	85306006	SRS1767352	SRP091570	NA	Urban	United States	North America
SRR4481776	68033378	SRS1767358	SRP091570	NA	Urban	United States	North America
SRR4481784	76420224	SRS1767362	SRP091570	NA	Urban	United States	North America
SRR4481801	79713392	SRS1767364	SRP091570	Diseased	Urban	United States	North America
SRR4481807	83953210	SRS1767365	SRP091570	NA	Urban	United States	North America
SRR4481809	94599798	SRS1767366	SRP091570	NA	Urban	United States	North America
SRR4481813	83792456	SRS1767367	SRP091570	NA	Urban	United States	North America
SRR4481820	69259776	SRS1767347	SRP091570	Diseased	Urban	United States	North America
SRR4783393	91991552	SRS1771173	SRP091570	Diseased	Urban	United States	North America
SRR4783406	81314204	SRS1771176	SRP091570	Diseased	Urban	United States	North America
SRR4783411	91266972	SRS1771179	SRP091570	NA	Urban	United States	North America
SRR4783413	76777616	SRS1771180	SRP091570	NA	Urban	United States	North America
SRR4783415	79471770	SRS1771175	SRP091570	Diseased	Urban	United States	North America
SRR4783426	89341304	SRS1771181	SRP091570	Diseased	Urban	United States	North America
SRR4783438	76683354	SRS1771177	SRP091570	NA	Urban	United States	North America
SRR4783440	73219340	SRS1771182	SRP091570	NA	Urban	United States	North America
SRR4783445	77928658	SRS1771178	SRP091570	NA	Urban	United States	North America
SRR4783448	74399676	SRS1767346	SRP091570	NA	Urban	United States	North America
SRR4783461	123970712	SRS1771184	SRP091570	Diseased	Urban	United States	North America
SRR4783488	80015292	SRS1771174	SRP091570	Diseased	Urban	United States	North America
SRR4783502	76479604	SRS1771188	SRP091570	NA	Urban	United States	North America
SRR4783504	77904056	SRS1771187	SRP091570	NA	Urban	United States	North America
SRR4783506	80495252	SRS1771186	SRP091570	Diseased	Urban	United States	North America
SRR4783510	82796772	SRS1771183	SRP091570	NA	Urban	United States	North America
SRR4783512	129644442	SRS1771189	SRP091570	NA	Urban	United States	North America
SRR4783522	72150344	SRS1771192	SRP091570	Diseased	Urban	United States	North America
SRR4783560	72704958	SRS1771196	SRP091570	Diseased	Urban	United States	North America

SRR4783567	90747132	SRS1771185	SRP091570	Diseased	Urban	United States	North America
SRR4783574	79089768	SRS1771194	SRP091570	NA	Urban	United States	North America
SRR4783589	73081966	SRS1756238	SRP091570	NA	Urban	United States	North America
SRR4783596	71575104	SRS1771193	SRP091570	Diseased	Urban	United States	North America
SRR4783607	80854954	SRS1771195	SRP091570	NA	Urban	United States	North America
SRR4783612	72282820	SRS1771197	SRP091570	NA	Urban	United States	North America
SRR4783629	74540252	SRS1771199	SRP091570	NA	Urban	United States	North America
SRR4783647	73007716	SRS1771200	SRP091570	NA	Urban	United States	North America
SRR4783650	77923254	SRS1771191	SRP091570	NA	Urban	United States	North America
SRR5024276	55924436	SRS1801062	SRP093449	NA	NA	United States	North America
SRR5024277	54720964	SRS1801064	SRP093449	NA	NA	United States	North America
SRR5024280	75989730	SRS1801067	SRP093449	NA	NA	United States	North America
SRR5024281	60243968	SRS1801068	SRP093449	NA	NA	United States	North America
SRR5024284	72059190	SRS1801071	SRP093449	NA	NA	United States	North America
SRR5024285	74987172	SRS1801073	SRP093449	NA	NA	United States	North America
SRR5032269	55708420	SRS1806014	SRP093506	Diseased	NA	China	Asia
SRR5032278	56513830	SRS1806023	SRP093506	Diseased	NA	China	Asia
SRR5032280	54448798	SRS1806024	SRP093506	Diseased	NA	China	Asia
SRR5032283	55726462	SRS1806028	SRP093506	Diseased	NA	China	Asia
SRR5032287	62862704	SRS1806032	SRP093506	Diseased	NA	China	Asia
SRR5032288	63844660	SRS1806033	SRP093506	Diseased	NA	China	Asia
SRR5032297	55700756	SRS1806042	SRP093506	Diseased	NA	China	Asia
SRR5032300	58176298	SRS1806045	SRP093506	Diseased	NA	China	Asia
SRR5032305	54224632	SRS1806050	SRP093506	Diseased	NA	China	Asia
SRR5032309	78248130	SRS1806054	SRP093506	Diseased	NA	China	Asia
SRR5032314	56414814	SRS1806059	SRP093506	Diseased	NA	China	Asia
SRR5032316	86032718	SRS1806061	SRP093506	Diseased	NA	China	Asia
SRR5032323	51822168	SRS1806068	SRP093506	Diseased	NA	China	Asia
SRR5032325	53512346	SRS1806070	SRP093506	Diseased	NA	China	Asia
SRR5032331	69011856	SRS1806076	SRP093506	Diseased	NA	China	Asia
SRR5032336	54802494	SRS1806081	SRP093506	Diseased	NA	China	Asia
SRR5032341	51256518	SRS1806086	SRP093506	Diseased	NA	China	Asia
SRR5032352	73197820	SRS1806097	SRP093506	Diseased	NA	China	Asia
SRR5032354	84925868	SRS1806099	SRP093506	Diseased	NA	China	Asia
SRR5032355	74097944	SRS1806100	SRP093506	Diseased	NA	China	Asia
SRR5032356	75015422	SRS1806101	SRP093506	Diseased	NA	China	Asia
SRR5056644	71429434	SRS1820192	SRP093965	Diseased	Urban	United States	North America
SRR5056645	63341554	SRS1820193	SRP093965	Diseased	Urban	United States	North America
SRR5056646	56105436	SRS1820194	SRP093965	Diseased	Urban	United States	North America
SRR5056647	67725164	SRS1820195	SRP093965	Diseased	Urban	United States	North America
SRR5056648	68644468	SRS1820196	SRP093965	Diseased	Urban	United States	North America
SRR5056649	60596486	SRS1820197	SRP093965	Diseased	Urban	United States	North America
SRR5056650	66703536	SRS1820198	SRP093965	Diseased	Urban	United States	North America
SRR5056652	68326692	SRS1820200	SRP093965	Diseased	Urban	United States	North America
SRR5056653	62295816	SRS1820201	SRP093965	Diseased	Urban	United States	North America
SRR5056654	66789522	SRS1820202	SRP093965	Diseased	Urban	United States	North America
SRR5056656	52649208	SRS1820204	SRP093965	Diseased	Urban	United States	North America
SRR5056659	60682602	SRS1820207	SRP093965	Diseased	Urban	United States	North America
SRR5056660	53500880	SRS1820208	SRP093965	Diseased	Urban	United States	North America
SRR5056661	66981254	SRS1820209	SRP093965	Diseased	Urban	United States	North America
SRR5056662	55660028	SRS1820210	SRP093965	Diseased	Urban	United States	North America
SRR5056664	60946506	SRS1820212	SRP093965	Diseased	Urban	United States	North America
SRR5056665	58964162	SRS1820213	SRP093965	Diseased	Urban	United States	North America
SRR5056666	61221730	SRS1820214	SRP093965	Diseased	Urban	United States	North America
SRR5056667	59511056	SRS1820215	SRP093965	Diseased	Urban	United States	North America
SRR5056668	66498866	SRS1820216	SRP093965	Diseased	Urban	United States	North America
SRR5056671	58266178	SRS1820219	SRP093965	Diseased	Urban	United States	North America
SRR5056673	50398306	SRS1820221	SRP093965	Diseased	Urban	United States	North America
SRR5056675	62280444	SRS1820223	SRP093965	Diseased	Urban	United States	North America
SRR5056676	55465910	SRS1820224	SRP093965	Diseased	Urban	United States	North America
SRR5056679	54548804	SRS1820227	SRP093965	Diseased	Urban	United States	North America
SRR5056681	78577872	SRS1820229	SRP093965	Diseased	Urban	United States	North America
SRR5056682	65169628	SRS1820230	SRP093965	Diseased	Urban	United States	North America
SRR5056683	68225020	SRS1820231	SRP093965	Diseased	Urban	United States	North America
SRR5056685	61561500	SRS1820233	SRP093965	Diseased	Urban	United States	North America
SRR5056686	59372536	SRS1820234	SRP093965	Diseased	Urban	United States	North America
SRR5056688	61015908	SRS1820236	SRP093965	Diseased	Urban	United States	North America
SRR5056689	59194712	SRS1820237	SRP093965	Diseased	Urban	United States	North America
SRR5056690	61655444	SRS1820238	SRP093965	Diseased	Urban	United States	North America
SRR5056693	57205048	SRS1820241	SRP093965	Diseased	Urban	United States	North America
SRR5056694	62150384	SRS1820242	SRP093965	Diseased	Urban	United States	North America
SRR5056695	64494080	SRS1820243	SRP093965	Diseased	Urban	United States	North America

SRR5056696	63190110	SRS1820244	SRP093965	Diseased	Urban	United States	North America
SRR5056698	53590334	SRS1820246	SRP093965	Diseased	Urban	United States	North America
SRR5056699	73949604	SRS1820247	SRP093965	Diseased	Urban	United States	North America
SRR5056700	75455030	SRS1820248	SRP093965	Diseased	Urban	United States	North America
SRR5056703	53871558	SRS1820251	SRP093965	Diseased	Urban	United States	North America
SRR5056704	54841318	SRS1820252	SRP093965	Diseased	Urban	United States	North America
SRR5056705	64936402	SRS1820253	SRP093965	Diseased	Urban	United States	North America
SRR5056706	65040090	SRS1820254	SRP093965	Diseased	Urban	United States	North America
SRR5056708	54448824	SRS1820256	SRP093965	Diseased	Urban	United States	North America
SRR5056711	53019936	SRS1820259	SRP093965	Diseased	Urban	United States	North America
SRR5056714	58875100	SRS1820240	SRP093965	Diseased	Urban	United States	North America
SRR5056715	65957854	SRS1820262	SRP093965	Diseased	Urban	United States	North America
SRR5056717	56354112	SRS1820264	SRP093965	Diseased	Urban	United States	North America
SRR5056718	61956958	SRS1820265	SRP093965	Diseased	Urban	United States	North America
SRR5056720	67137272	SRS1820199	SRP093965	Diseased	Urban	United States	North America
SRR5056721	63598306	SRS1820267	SRP093965	Diseased	Urban	United States	North America
SRR5056724	58823626	SRS1820270	SRP093965	Diseased	Urban	United States	North America
SRR5056725	62858466	SRS1820271	SRP093965	Diseased	Urban	United States	North America
SRR5056728	61562132	SRS1820274	SRP093965	Diseased	Urban	United States	North America
SRR5056730	63872306	SRS1820276	SRP093965	Diseased	Urban	United States	North America
SRR5056731	71940506	SRS1820277	SRP093965	Diseased	Urban	United States	North America
SRR5056733	54448886	SRS1820279	SRP093965	Diseased	Urban	United States	North America
SRR5056735	64180506	SRS1820281	SRP093965	Diseased	Urban	United States	North America
SRR5056738	61972518	SRS1820283	SRP093965	Diseased	Urban	United States	North America
SRR5056740	60768094	SRS1820261	SRP093965	Diseased	Urban	United States	North America
SRR5056743	62896212	SRS1820287	SRP093965	Diseased	Urban	United States	North America
SRR5056745	62410226	SRS1820289	SRP093965	Diseased	Urban	United States	North America
SRR5056748	70183180	SRS1820292	SRP093965	Diseased	Urban	United States	North America
SRR5056749	64085412	SRS1820293	SRP093965	Diseased	Urban	United States	North America
SRR5056750	60592702	SRS1820296	SRP093965	Diseased	Urban	United States	North America
SRR5056756	59697078	SRS1820298	SRP093965	Diseased	Urban	United States	North America
SRR5056757	81357576	SRS1820299	SRP093965	Diseased	Urban	United States	North America
SRR5056759	65586846	SRS1820291	SRP093965	Diseased	Urban	United States	North America
SRR5056760	54805830	SRS1820301	SRP093965	Diseased	Urban	United States	North America
SRR5056761	57535240	SRS1820302	SRP093965	Diseased	Urban	United States	North America
SRR5056763	59304542	SRS1820303	SRP093965	Diseased	Urban	United States	North America
SRR5056770	59611686	SRS1820310	SRP093965	Diseased	Urban	United States	North America
SRR5056772	60839326	SRS1820311	SRP093965	Diseased	Urban	United States	North America
SRR5056774	67553498	SRS1820314	SRP093965	Diseased	Urban	United States	North America
SRR5056775	55306028	SRS1820315	SRP093965	Diseased	Urban	United States	North America
SRR5056777	66928856	SRS1820317	SRP093965	Diseased	Urban	United States	North America
SRR5056779	63851470	SRS1820318	SRP093965	Diseased	Urban	United States	North America
SRR5056780	61741572	SRS1820319	SRP093965	Diseased	Urban	United States	North America
SRR5056782	52140158	SRS1820320	SRP093965	Diseased	Urban	United States	North America
SRR5056784	57231324	SRS1820235	SRP093965	Diseased	Urban	United States	North America
SRR5056786	72194650	SRS1820323	SRP093965	Diseased	Urban	United States	North America
SRR5056788	53749100	SRS1820220	SRP093965	Diseased	Urban	United States	North America
SRR5056789	55972550	SRS1820325	SRP093965	Diseased	Urban	United States	North America
SRR5056790	58246022	SRS1820326	SRP093965	Diseased	Urban	United States	North America
SRR5056791	55353450	SRS1820327	SRP093965	Diseased	Urban	United States	North America
SRR5056792	53264302	SRS1820328	SRP093965	Diseased	Urban	United States	North America
SRR5056793	54617710	SRS1820330	SRP093965	Diseased	Urban	United States	North America
SRR5056794	59811104	SRS1820329	SRP093965	Diseased	Urban	United States	North America
SRR5056795	53400848	SRS1820331	SRP093965	Diseased	Urban	United States	North America
SRR5056796	76281246	SRS1820332	SRP093965	Diseased	Urban	United States	North America
SRR5056797	66704978	SRS1820333	SRP093965	Diseased	Urban	United States	North America
SRR5056798	56633056	SRS1820334	SRP093965	Diseased	Urban	United States	North America
SRR5056801	55177842	SRS1820336	SRP093965	Diseased	Urban	United States	North America
SRR5056807	54005896	SRS1820343	SRP093965	Diseased	Urban	United States	North America
SRR5056808	82871744	SRS1820342	SRP093965	Diseased	Urban	United States	North America
SRR5056810	66505568	SRS1820344	SRP093965	Diseased	Urban	United States	North America
SRR5056813	62156776	SRS1820347	SRP093965	Diseased	Urban	United States	North America
SRR5056814	77747520	SRS1820348	SRP093965	Diseased	Urban	United States	North America
SRR5056815	60065182	SRS1820349	SRP093965	Diseased	Urban	United States	North America
SRR5056816	57404114	SRS1820306	SRP093965	Diseased	Urban	United States	North America
SRR5056817	52647398	SRS1820340	SRP093965	Diseased	Urban	United States	North America
SRR5056818	73964156	SRS1820350	SRP093965	Diseased	Urban	United States	North America
SRR5056819	61424784	SRS1820232	SRP093965	Diseased	Urban	United States	North America
SRR5056823	57608040	SRS1820307	SRP093965	Diseased	Urban	United States	North America
SRR5056824	60024280	SRS1820353	SRP093965	Diseased	Urban	United States	North America
SRR5056825	66847284	SRS1820354	SRP093965	Diseased	Urban	United States	North America
SRR5056826	64639810	SRS1820355	SRP093965	Diseased	Urban	United States	North America

SRR5056827	55830868	SRS1820356	SRP093965	Diseased	Urban	United States	North America
SRR5056828	68290604	SRS1820357	SRP093965	Diseased	Urban	United States	North America
SRR5056829	64057118	SRS1820358	SRP093965	Diseased	Urban	United States	North America
SRR5056831	116452386	SRS1820360	SRP093965	Diseased	Urban	United States	North America
SRR5056832	54012434	SRS1820312	SRP093965	Diseased	Urban	United States	North America
SRR5056833	60455690	SRS1820361	SRP093965	Diseased	Urban	United States	North America
SRR5056834	52226558	SRS1820362	SRP093965	Diseased	Urban	United States	North America
SRR5056836	57688682	SRS1820364	SRP093965	Diseased	Urban	United States	North America
SRR5056837	53130382	SRS1820365	SRP093965	Diseased	Urban	United States	North America
SRR5056838	66032288	SRS1820366	SRP093965	Diseased	Urban	United States	North America
SRR5056839	66921994	SRS1820280	SRP093965	Diseased	Urban	United States	North America
SRR5056841	69569384	SRS1820368	SRP093965	Diseased	Urban	United States	North America
SRR5056843	68607706	SRS1820369	SRP093965	Diseased	Urban	United States	North America
SRR5056844	61914108	SRS1820371	SRP093965	Diseased	Urban	United States	North America
SRR5056845	63253242	SRS1820372	SRP093965	Diseased	Urban	United States	North America
SRR5056846	58901578	SRS1820374	SRP093965	Diseased	Urban	United States	North America
SRR5056847	56663232	SRS1820373	SRP093965	Diseased	Urban	United States	North America
SRR5056851	76258188	SRS1820378	SRP093965	Diseased	Urban	United States	North America
SRR5056852	67784814	SRS1820376	SRP093965	Diseased	Urban	United States	North America
SRR5056853	60978796	SRS1820379	SRP093965	Diseased	Urban	United States	North America
SRR5056854	69485234	SRS1820381	SRP093965	Diseased	Urban	United States	North America
SRR5056856	56901464	SRS1820382	SRP093965	Diseased	Urban	United States	North America
SRR5056857	71946560	SRS1820383	SRP093965	Diseased	Urban	United States	North America
SRR5056858	80323094	SRS1820384	SRP093965	Diseased	Urban	United States	North America
SRR5056859	63011430	SRS1820385	SRP093965	Diseased	Urban	United States	North America
SRR5056860	63487324	SRS1820386	SRP093965	Diseased	Urban	United States	North America
SRR5056861	57023700	SRS1820387	SRP093965	Diseased	Urban	United States	North America
SRR5056864	54601218	SRS1820389	SRP093965	Diseased	Urban	United States	North America
SRR5056866	59769880	SRS1820391	SRP093965	Diseased	Urban	United States	North America
SRR5056867	58760912	SRS1820392	SRP093965	Diseased	Urban	United States	North America
SRR5056868	56772054	SRS1820393	SRP093965	Diseased	Urban	United States	North America
SRR5056870	56422870	SRS1820269	SRP093965	Diseased	Urban	United States	North America
SRR5056873	58526372	SRS1820396	SRP093965	Diseased	Urban	United States	North America
SRR5056874	55855326	SRS1820397	SRP093965	Diseased	Urban	United States	North America
SRR5056875	71913006	SRS1820263	SRP093965	Diseased	Urban	United States	North America
SRR5056876	59265636	SRS1820398	SRP093965	Diseased	Urban	United States	North America
SRR5056877	75622816	SRS1820399	SRP093965	Diseased	Urban	United States	North America
SRR5056880	66215570	SRS1820400	SRP093965	Diseased	Urban	United States	North America
SRR5056883	58652380	SRS1820403	SRP093965	Diseased	Urban	United States	North America
SRR5056884	69908802	SRS1820404	SRP093965	Diseased	Urban	United States	North America
SRR5056886	71275618	SRS1820406	SRP093965	Diseased	Urban	United States	North America
SRR5056887	54309198	SRS1820407	SRP093965	Diseased	Urban	United States	North America
SRR5056888	78696986	SRS1820408	SRP093965	Diseased	Urban	United States	North America
SRR5056889	62761562	SRS1820409	SRP093965	Diseased	Urban	United States	North America
SRR5056890	76073130	SRS1820370	SRP093965	Diseased	Urban	United States	North America
SRR5056891	78547220	SRS1820410	SRP093965	Diseased	Urban	United States	North America
SRR5056892	51823954	SRS1820401	SRP093965	Diseased	Urban	United States	North America
SRR5056893	68108672	SRS1820411	SRP093965	Diseased	Urban	United States	North America
SRR5056894	59632542	SRS1820412	SRP093965	Diseased	Urban	United States	North America
SRR5056895	57528258	SRS1820413	SRP093965	Diseased	Urban	United States	North America
SRR5056896	53482158	SRS1820414	SRP093965	Diseased	Urban	United States	North America
SRR5056898	56285984	SRS1820416	SRP093965	Diseased	Urban	United States	North America
SRR5056900	64076458	SRS1820418	SRP093965	Diseased	Urban	United States	North America
SRR5056901	59112392	SRS1820278	SRP093965	Diseased	Urban	United States	North America
SRR5056904	63525506	SRS1820420	SRP093965	Diseased	Urban	United States	North America
SRR5056905	60869936	SRS1820421	SRP093965	Diseased	Urban	United States	North America
SRR5056906	58407674	SRS1820422	SRP093965	Diseased	Urban	United States	North America
SRR5056907	54906332	SRS1820423	SRP093965	Diseased	Urban	United States	North America
SRR5056915	52381500	SRS1820428	SRP093965	Diseased	Urban	United States	North America
SRR5056916	58048218	SRS1820429	SRP093965	Diseased	Urban	United States	North America
SRR5056917	65608686	SRS1820430	SRP093965	Diseased	Urban	United States	North America
SRR5056918	58425076	SRS1820431	SRP093965	Diseased	Urban	United States	North America
SRR5056919	77510042	SRS1820432	SRP093965	Diseased	Urban	United States	North America
SRR5056920	70771860	SRS1820211	SRP093965	Diseased	Urban	United States	North America
SRR5056921	68889512	SRS1820433	SRP093965	Diseased	Urban	United States	North America
SRR5056923	57352590	SRS1820434	SRP093965	Diseased	Urban	United States	North America
SRR5056926	50623428	SRS1820436	SRP093965	Diseased	Urban	United States	North America
SRR5056929	59628886	SRS1820439	SRP093965	Diseased	Urban	United States	North America
SRR5056930	58789840	SRS1820380	SRP093965	Diseased	Urban	United States	North America
SRR5056931	55437262	SRS1820321	SRP093965	Diseased	Urban	United States	North America
SRR5056932	72131730	SRS1820440	SRP093965	Diseased	Urban	United States	North America
SRR5056934	64654078	SRS1820442	SRP093965	Diseased	Urban	United States	North America

SRR5056935	74117348	SRS1820441	SRP093965	Diseased	Urban	United States	North America
SRR5056937	61784316	SRS1820228	SRP093965	Diseased	Urban	United States	North America
SRR5056938	61938664	SRS1820443	SRP093965	Diseased	Urban	United States	North America
SRR5056942	57442850	SRS1820447	SRP093965	Diseased	Urban	United States	North America
SRR5056943	70963090	SRS1820448	SRP093965	Diseased	Urban	United States	North America
SRR5056944	65040138	SRS1820449	SRP093965	Diseased	Urban	United States	North America
SRR5056946	60802272	SRS1820450	SRP093965	Diseased	Urban	United States	North America
SRR5056948	70846600	SRS1820452	SRP093965	Diseased	Urban	United States	North America
SRR5056950	53824174	SRS1820453	SRP093965	Diseased	Urban	United States	North America
SRR5056951	54698862	SRS1820454	SRP093965	Diseased	Urban	United States	North America
SRR5056952	59746820	SRS1820455	SRP093965	Diseased	Urban	United States	North America
SRR5056953	68022546	SRS1820456	SRP093965	Diseased	Urban	United States	North America
SRR5056955	64539118	SRS1820458	SRP093965	Diseased	Urban	United States	North America
SRR5056960	55750084	SRS1820461	SRP093965	Diseased	Urban	United States	North America
SRR5056961	63230140	SRS1820462	SRP093965	Diseased	Urban	United States	North America
SRR5056967	52485604	SRS1820465	SRP093965	Diseased	Urban	United States	North America
SRR5056968	67790148	SRS1820466	SRP093965	Diseased	Urban	United States	North America
SRR5056969	60513206	SRS1820467	SRP093965	Diseased	Urban	United States	North America
SRR5056970	69848688	SRS1820468	SRP093965	Diseased	Urban	United States	North America
SRR5056973	70691610	SRS1820470	SRP093965	Diseased	Urban	United States	North America
SRR5056975	58364960	SRS1820471	SRP093965	Diseased	Urban	United States	North America
SRR5056976	58330264	SRS1820472	SRP093965	Diseased	Urban	United States	North America
SRR5056979	65607894	SRS1820473	SRP093965	Diseased	Urban	United States	North America
SRR5056980	58958908	SRS1820322	SRP093965	Diseased	Urban	United States	North America
SRR5056982	64753784	SRS1820337	SRP093965	Diseased	Urban	United States	North America
SRR5056984	56926378	SRS1820475	SRP093965	Diseased	Urban	United States	North America
SRR5056987	59888032	SRS1820477	SRP093965	Diseased	Urban	United States	North America
SRR5056988	66431968	SRS1820359	SRP093965	Diseased	Urban	United States	North America
SRR5056989	62690854	SRS1820363	SRP093965	Diseased	Urban	United States	North America
SRR5056990	58599972	SRS1820478	SRP093965	Diseased	Urban	United States	North America
SRR5056991	66816374	SRS1820425	SRP093965	Diseased	Urban	United States	North America
SRR5056992	56574936	SRS1820479	SRP093965	Diseased	Urban	United States	North America
SRR5056994	54479152	SRS1820481	SRP093965	Diseased	Urban	United States	North America
SRR5056996	72959268	SRS1820482	SRP093965	Diseased	Urban	United States	North America
SRR5056999	67439436	SRS1820484	SRP093965	Diseased	Urban	United States	North America
SRR5057000	69663588	SRS1820357	SRP093965	Diseased	Urban	United States	North America
SRR5057007	60056048	SRS1820487	SRP093965	Diseased	Urban	United States	North America
SRR5057009	58417942	SRS1820488	SRP093965	Diseased	Urban	United States	North America
SRR5057012	59170922	SRS1820490	SRP093965	Diseased	Urban	United States	North America
SRR5057013	62065714	SRS1820491	SRP093965	Diseased	Urban	United States	North America
SRR5057014	236988332	SRS1820492	SRP093965	Diseased	Urban	United States	North America
SRR5057015	66893682	SRS1820493	SRP093965	Diseased	Urban	United States	North America
SRR5057019	55867564	SRS1820495	SRP093965	Diseased	Urban	United States	North America
SRR5057020	65121042	SRS1820496	SRP093965	Diseased	Urban	United States	North America
SRR5057021	56884162	SRS1820497	SRP093965	Diseased	Urban	United States	North America
SRR5057024	52031236	SRS1820499	SRP093965	Diseased	Urban	United States	North America
SRR5057027	51071230	SRS1820451	SRP093965	Diseased	Urban	United States	North America
SRR5057030	74299924	SRS1820341	SRP093965	Diseased	Urban	United States	North America
SRR5057032	51764674	SRS1820503	SRP093965	Diseased	Urban	United States	North America
SRR5057033	78988882	SRS1820504	SRP093965	Diseased	Urban	United States	North America
SRR5057034	69513914	SRS1820505	SRP093965	Diseased	Urban	United States	North America
SRR5057036	60994042	SRS1820507	SRP093965	Diseased	Urban	United States	North America
SRR5057037	69044272	SRS1820255	SRP093965	Diseased	Urban	United States	North America
SRR5057038	57614296	SRS1820508	SRP093965	Diseased	Urban	United States	North America
SRR5057039	62741662	SRS1820510	SRP093965	Diseased	Urban	United States	North America
SRR5057040	63770318	SRS1820272	SRP093965	Diseased	Urban	United States	North America
SRR5057041	57407538	SRS1820509	SRP093965	Diseased	Urban	United States	North America
SRR5057042	72688280	SRS1820511	SRP093965	Diseased	Urban	United States	North America
SRR5057043	53876514	SRS1820512	SRP093965	Diseased	Urban	United States	North America
SRR5057044	59073376	SRS1820513	SRP093965	Diseased	Urban	United States	North America
SRR5057045	63846450	SRS1820516	SRP093965	Diseased	Urban	United States	North America
SRR5057046	53445366	SRS1820426	SRP093965	Diseased	Urban	United States	North America
SRR5057047	60483922	SRS1820514	SRP093965	Diseased	Urban	United States	North America
SRR5057048	61962182	SRS1820515	SRP093965	Diseased	Urban	United States	North America
SRR5057049	52116472	SRS1820517	SRP093965	Diseased	Urban	United States	North America
SRR5057052	59179410	SRS1820519	SRP093965	Diseased	Urban	United States	North America
SRR5057054	54583528	SRS1820521	SRP093965	Diseased	Urban	United States	North America
SRR5057059	60248684	SRS1820304	SRP093965	Diseased	Urban	United States	North America
SRR5057061	51197910	SRS1820525	SRP093965	Diseased	Urban	United States	North America
SRR5057062	51136540	SRS1820526	SRP093965	Diseased	Urban	United States	North America
SRR5057065	55226908	SRS1820217	SRP093965	Diseased	Urban	United States	North America
SRR5057067	53291972	SRS1820480	SRP093965	Diseased	Urban	United States	North America

SRR5057068	64847778	SRS1820528	SRP093965	Diseased	Urban	United States	North America
SRR5057070	62188520	SRS1820367	SRP093965	Diseased	Urban	United States	North America
SRR5057073	67800924	SRS1820532	SRP093965	Diseased	Urban	United States	North America
SRR5057074	59738328	SRS1820225	SRP093965	Diseased	Urban	United States	North America
SRR5057076	53909278	SRS1820469	SRP093965	Diseased	Urban	United States	North America
SRR5057080	71596520	SRS1820388	SRP093965	Diseased	Urban	United States	North America
SRR5057082	59364428	SRS1820534	SRP093965	Diseased	Urban	United States	North America
SRR5057083	54296718	SRS1820536	SRP093965	Diseased	Urban	United States	North America
SRR5057085	67570768	SRS1820460	SRP093965	Diseased	Urban	United States	North America
SRR5057086	53954444	SRS1820537	SRP093965	Diseased	Urban	United States	North America
SRR5057088	58711532	SRS1820494	SRP093965	Diseased	Urban	United States	North America
SRR5057089	79139386	SRS1820538	SRP093965	Diseased	Urban	United States	North America
SRR5057090	75608310	SRS1820424	SRP093965	Diseased	Urban	United States	North America
SRR5057091	61067240	SRS1820539	SRP093965	Diseased	Urban	United States	North America
SRR5057092	55790064	SRS1820540	SRP093965	Diseased	Urban	United States	North America
SRR5057094	58263854	SRS1820542	SRP093965	Diseased	Urban	United States	North America
SRR5057097	63665366	SRS1820544	SRP093965	Diseased	Urban	United States	North America
SRR5057100	60855576	SRS1820222	SRP093965	Diseased	Urban	United States	North America
SRR5057102	69673006	SRS1820547	SRP093965	Diseased	Urban	United States	North America
SRR5057106	57791330	SRS1820550	SRP093965	Diseased	Urban	United States	North America
SRR5057110	58756360	SRS1820527	SRP093965	Diseased	Urban	United States	North America
SRR5057112	70922420	SRS1820555	SRP093965	Diseased	Urban	United States	North America
SRR5057113	62689328	SRS1820557	SRP093965	Diseased	Urban	United States	North America
SRR5057115	58533322	SRS1820266	SRP093965	Diseased	Urban	United States	North America
SRR5057116	73641030	SRS1820556	SRP093965	Diseased	Urban	United States	North America
SRR5057119	64599450	SRS1820301	SRP093965	Diseased	Urban	United States	North America
SRR5057121	67830650	SRS1820558	SRP093965	Diseased	Urban	United States	North America
SRR5057124	63825558	SRS1820377	SRP093965	Diseased	Urban	United States	North America
SRR5057125	78791286	SRS1820560	SRP093965	Diseased	Urban	United States	North America
SRR5057126	69655314	SRS1820337	SRP093965	Diseased	Urban	United States	North America
SRR5057127	51490178	SRS1820562	SRP093965	Diseased	Urban	United States	North America
SRR5057128	64989868	SRS1820563	SRP093965	Diseased	Urban	United States	North America
SRR5077148	69745240	SRS1820285	SRP093965	Diseased	Urban	United States	North America
SRR5091457	61936772	SRS1846747	SRP094805	NA	NA	China	Asia
SRR5091461	91300540	SRS1846751	SRP094805	NA	NA	China	Asia
SRR5091464	83953552	SRS1846753	SRP094805	NA	NA	China	Asia
SRR5091467	86934574	SRS1846757	SRP094805	NA	NA	China	Asia
SRR5091473	112553632	SRS1846763	SRP094805	NA	NA	China	Asia
SRR5091477	56250162	SRS1846767	SRP094805	NA	NA	China	Asia
SRR5091478	279718300	SRS1846768	SRP094805	NA	NA	China	Asia
SRR5091482	72243090	SRS1846771	SRP094805	NA	NA	China	Asia
SRR5091484	53735974	SRS1846774	SRP094805	NA	NA	China	Asia
SRR5091485	65203498	SRS1846775	SRP094805	NA	NA	China	Asia
SRR5091486	64565996	SRS1846776	SRP094805	NA	NA	China	Asia
SRR5091491	82253840	SRS1846781	SRP094805	NA	NA	China	Asia
SRR5091492	56384622	SRS1846782	SRP094805	NA	NA	China	Asia
SRR5091496	51309414	SRS1846786	SRP094805	NA	NA	China	Asia
SRR5091497	63812742	SRS1846787	SRP094805	NA	NA	China	Asia
SRR5091499	50537684	SRS1846790	SRP094805	NA	NA	China	Asia
SRR5091500	61454320	SRS1846789	SRP094805	NA	NA	China	Asia
SRR5091505	62693714	SRS1846794	SRP094805	NA	NA	China	Asia
SRR5091508	50807918	SRS1846798	SRP094805	NA	NA	China	Asia
SRR5091509	71163418	SRS1846799	SRP094805	NA	NA	China	Asia
SRR5091510	64061772	SRS1846801	SRP094805	NA	NA	China	Asia
SRR5091513	59481866	SRS1846803	SRP094805	NA	NA	China	Asia
SRR5091514	78434636	SRS1846804	SRP094805	NA	NA	China	Asia
SRR5091516	61888458	SRS1846806	SRP094805	NA	NA	China	Asia
SRR5091518	62607876	SRS1846808	SRP094805	NA	NA	China	Asia
SRR5091521	84016312	SRS1846811	SRP094805	NA	NA	China	Asia
SRR5091523	72131332	SRS1846813	SRP094805	NA	NA	China	Asia
SRR5091526	66817384	SRS1846815	SRP094805	NA	NA	China	Asia
SRR5091528	55343226	SRS1846819	SRP094805	NA	NA	China	Asia
SRR5091529	55718172	SRS1846818	SRP094805	NA	NA	China	Asia
SRR5091531	269118492	SRS1846822	SRP094805	NA	NA	China	Asia
SRR5091541	70284244	SRS1846831	SRP094805	NA	NA	China	Asia
SRR5091544	80089088	SRS1846834	SRP094805	NA	NA	China	Asia
SRR5091546	62957452	SRS1846836	SRP094805	NA	NA	China	Asia
SRR5091547	50537010	SRS1846838	SRP094805	NA	NA	China	Asia
SRR5091555	64583632	SRS1846845	SRP094805	NA	NA	China	Asia
SRR5091556	85957104	SRS1846846	SRP094805	NA	NA	China	Asia
SRR5091559	61462538	SRS1846849	SRP094805	NA	NA	China	Asia
SRR5091562	72811476	SRS1846852	SRP094805	NA	NA	China	Asia

SRR5091564	129194422	SRS1846854	SRP094805	NA	NA	China	Asia
SRR5091568	63039254	SRS1846858	SRP094805	NA	NA	China	Asia
SRR5091569	55283840	SRS1846859	SRP094805	NA	NA	China	Asia
SRR5091570	54296392	SRS1846860	SRP094805	NA	NA	China	Asia
SRR5091574	61152170	SRS1846864	SRP094805	NA	NA	China	Asia
SRR5091576	52384860	SRS1846866	SRP094805	NA	NA	China	Asia
SRR5091580	93454420	SRS1846870	SRP094805	NA	NA	China	Asia
SRR5091581	52489080	SRS1846872	SRP094805	NA	NA	China	Asia
SRR5091591	70078422	SRS1846881	SRP094805	NA	NA	China	Asia
SRR5091593	50521456	SRS1846883	SRP094805	NA	NA	China	Asia
SRR5091596	62656886	SRS1846886	SRP094805	NA	NA	China	Asia
SRR5091598	56222858	SRS1846888	SRP094805	NA	NA	China	Asia
SRR5091600	60065466	SRS1846889	SRP094805	NA	NA	China	Asia
SRR5091601	50264204	SRS1846891	SRP094805	NA	NA	China	Asia
SRR5091605	50211280	SRS1846895	SRP094805	NA	NA	China	Asia
SRR5091609	58392012	SRS1846899	SRP094805	NA	NA	China	Asia
SRR5091611	87288030	SRS1846901	SRP094805	NA	NA	China	Asia
SRR5091618	56314208	SRS1846908	SRP094805	NA	NA	China	Asia
SRR5106285	60697988	SRS1858594	SRP095060	Healthy	NA	China	Asia
SRR5106292	81803054	SRS1858601	SRP095060	Healthy	NA	China	Asia
SRR5106295	52173936	SRS1858604	SRP095060	Healthy	NA	China	Asia
SRR5106304	55947490	SRS1858613	SRP095060	Healthy	NA	China	Asia
SRR5106315	51684338	SRS1858624	SRP095060	Healthy	NA	China	Asia
SRR5106317	122143650	SRS1858626	SRP095060	Healthy	NA	China	Asia
SRR5106319	54763724	SRS1858628	SRP095060	Healthy	NA	China	Asia
SRR5106320	67968636	SRS1858629	SRP095060	Healthy	NA	China	Asia
SRR5106335	50376602	SRS1858644	SRP095060	Healthy	NA	China	Asia
SRR5106392	53038716	SRS1858663	SRP095060	Healthy	NA	China	Asia
SRR5106393	80840614	SRS1858664	SRP095060	Healthy	NA	China	Asia
SRR5106457	61979448	SRS1858681	SRP095060	Healthy	NA	China	Asia
SRR5106464	60776176	SRS1858688	SRP095060	Healthy	NA	China	Asia
SRR5106468	52492596	SRS1858692	SRP095060	Healthy	NA	China	Asia
SRR5106479	55657180	SRS1858703	SRP095060	Healthy	NA	China	Asia
SRR5106491	53829138	SRS1858715	SRP095060	Healthy	NA	China	Asia
SRR5106493	55082076	SRS1858717	SRP095060	Healthy	NA	China	Asia
SRR5106494	50323752	SRS1858718	SRP095060	Healthy	NA	China	Asia
SRR5106497	78480574	SRS1858721	SRP095060	Healthy	NA	China	Asia
SRR5127422	59941528	SRS1876289	SRP095580	Healthy	Urban	Netherlands	Europe
SRR5127502	61617930	SRS1876369	SRP095580	Healthy	Urban	Netherlands	Europe
SRR5127544	62721388	SRS1876411	SRP095580	Healthy	Urban	Netherlands	Europe
SRR5127571	51875844	SRS1876438	SRP095580	Healthy	Urban	Netherlands	Europe
SRR5127608	54950162	SRS1876475	SRP095580	Healthy	Urban	Netherlands	Europe
SRR5127631	63166278	SRS1876498	SRP095580	Healthy	Urban	Netherlands	Europe
SRR5127662	59529614	SRS1876529	SRP095580	Healthy	Urban	Netherlands	Europe
SRR5127666	62756934	SRS1876533	SRP095580	Healthy	Urban	Netherlands	Europe
SRR5127721	52404630	SRS1876588	SRP095580	Healthy	Urban	Netherlands	Europe
SRR5127724	61937484	SRS1876589	SRP095580	Healthy	Urban	Netherlands	Europe
SRR5127734	68347052	SRS1876601	SRP095580	Healthy	Urban	Netherlands	Europe
SRR5127754	50746656	SRS1876621	SRP095580	Healthy	Urban	Netherlands	Europe
SRR5127833	51796766	SRS1876700	SRP095580	Healthy	Urban	Netherlands	Europe
SRR5127853	58112738	SRS1876720	SRP095580	Healthy	Urban	Netherlands	Europe
SRR6028221	71673304	SRS1905887	SRP096283	Healthy	Urban	United States	North America
SRR6028495	60801050	SRS1905773	SRP096283	Healthy	Urban	United States	North America
SRR6038200	53999202	SRS1913090	SRP096283	Healthy	Urban	United States	North America
SRR6038223	54769448	SRS1912758	SRP096283	Healthy	Urban	United States	North America
SRR6038243	67156898	SRS1913072	SRP096283	Healthy	Urban	United States	North America
SRR6038266	52536342	SRS1912771	SRP096283	Healthy	Urban	United States	North America
SRR6038359	56244250	SRS1912818	SRP096283	Healthy	Urban	United States	North America
SRR6038399	84615180	SRS1912991	SRP096283	Healthy	Urban	United States	North America
SRR6038442	70893134	SRS1913012	SRP096283	Healthy	Urban	United States	North America
SRR6038472	53814756	SRS1912760	SRP096283	Healthy	Urban	United States	North America
SRR6038504	58835478	SRS1913083	SRP096283	Healthy	Urban	United States	North America
SRR6038526	101146224	SRS1913018	SRP096283	Healthy	Urban	United States	North America
SRR5275395	88454770	SRS1993527	SRP100446	Diseased	Urban	United States	North America
SRR5275397	71134850	SRS1993528	SRP100446	Diseased	Urban	United States	North America
SRR5275399	57023806	SRS1993530	SRP100446	Diseased	Urban	United States	North America
SRR5275403	51751594	SRS1993535	SRP100446	Diseased	Urban	United States	North America
SRR5275405	50991438	SRS1993537	SRP100446	Diseased	Urban	United States	North America
SRR5275409	111356470	SRS1993541	SRP100446	Diseased	Urban	United States	North America
SRR5275410	53392086	SRS1993543	SRP100446	Diseased	Urban	United States	North America
SRR5275412	51049826	SRS1993544	SRP100446	Diseased	Urban	United States	North America
SRR5275413	78805898	SRS1993545	SRP100446	Diseased	Urban	United States	North America

SRR5275414	78718356	SRS1993546	SRP100446	Diseased	Urban	United States	North America
SRR5275421	73959548	SRS1993548	SRP100446	Diseased	Urban	United States	North America
SRR5275424	60846928	SRS1993552	SRP100446	Diseased	Urban	United States	North America
SRR5275426	70125880	SRS1993553	SRP100446	Diseased	Urban	United States	North America
SRR5275430	59606474	SRS1993559	SRP100446	Diseased	Urban	United States	North America
SRR5275431	90679204	SRS1993557	SRP100446	Diseased	Urban	United States	North America
SRR5275432	55796296	SRS1993558	SRP100446	Diseased	Urban	United States	North America
SRR5275433	98578946	SRS1993561	SRP100446	Diseased	Urban	United States	North America
SRR5275435	56268970	SRS1993562	SRP100446	Diseased	Urban	United States	North America
SRR5275436	59389556	SRS1993563	SRP100446	Diseased	Urban	United States	North America
SRR5275437	68562686	SRS1993564	SRP100446	Diseased	Urban	United States	North America
SRR5275438	67243376	SRS1993565	SRP100446	Diseased	Urban	United States	North America
SRR5275444	61144210	SRS1993570	SRP100446	Diseased	Urban	United States	North America
SRR5275445	78282718	SRS1993573	SRP100446	Diseased	Urban	United States	North America
SRR5275447	73429434	SRS1993574	SRP100446	Diseased	Urban	United States	North America
SRR5275448	72877842	SRS1993575	SRP100446	Diseased	Urban	United States	North America
SRR5275452	68946890	SRS1993581	SRP100446	Diseased	Urban	United States	North America
SRR5275453	58694970	SRS1993578	SRP100446	Diseased	Urban	United States	North America
SRR5275455	179613510	SRS1993582	SRP100446	Diseased	Urban	United States	North America
SRR5275456	176104478	SRS1993583	SRP100446	Diseased	Urban	United States	North America
SRR5275457	76089210	SRS1993585	SRP100446	Diseased	Urban	United States	North America
SRR5275458	74411244	SRS1993587	SRP100446	Diseased	Urban	United States	North America
SRR5275465	55951308	SRS1994010	SRP100446	Diseased	Urban	United States	North America
SRR5275466	63345528	SRS1994013	SRP100446	Diseased	Urban	United States	North America
SRR5275468	62011188	SRS1994015	SRP100446	Diseased	Urban	United States	North America
SRR5275472	56083298	SRS1994020	SRP100446	Diseased	Urban	United States	North America
SRR5275473	62151150	SRS1994019	SRP100446	Diseased	Urban	United States	North America
SRR5275475	66364556	SRS1994023	SRP100446	Diseased	Urban	United States	North America
SRR5275476	80656876	SRS1994024	SRP100446	Diseased	Urban	United States	North America
SRR5275477	75077134	SRS1994025	SRP100446	Diseased	NA	United States	North America
SRR5275478	56605808	SRS1994022	SRP100446	Diseased	Urban	United States	North America
SRR5275480	59808492	SRS1994027	SRP100446	Diseased	Urban	United States	North America
SRR5275481	121489588	SRS1994028	SRP100446	Diseased	Urban	United States	North America
SRR5275482	290630554	SRS1994029	SRP100446	Diseased	NA	United States	North America
SRR5558154	58333324	SRS1996016	SRP100518	Diseased	Urban	Spain	Europe
SRR5558155	53326940	SRS1996016	SRP100518	Diseased	Urban	Spain	Europe
SRR5558232	58367112	SRS1996016	SRP100518	Diseased	Urban	Spain	Europe
SRR5558303	59821376	SRS1996016	SRP100518	Diseased	Urban	Spain	Europe
SRR5558304	60455120	SRS1996016	SRP100518	Diseased	Urban	Spain	Europe
SRR5558406	57496434	SRS1996016	SRP100518	Diseased	Urban	Spain	Europe
SRR5579956	58548020	SRS1996016	SRP100518	Diseased	Urban	Spain	Europe
SRR5580260	63171638	SRS1996016	SRP100518	Diseased	Urban	Spain	Europe
SRR5580272	63497894	SRS1996016	SRP100518	Diseased	Urban	Spain	Europe
SRR5580316	61225762	SRS1996016	SRP100518	Diseased	Urban	Spain	Europe
SRR5713943	80377570	SRS2296668	SRP100518	Diseased	Urban	Sweden	Europe
SRR5279217	50136012	SRS1996996	SRP100575	Diseased	NA	China	Asia
SRR5279222	56866368	SRS1997001	SRP100575	Diseased	NA	China	Asia
SRR5279223	55981960	SRS1997002	SRP100575	Diseased	NA	China	Asia
SRR5279225	50930494	SRS1997004	SRP100575	Diseased	NA	China	Asia
SRR5279226	61108312	SRS1997005	SRP100575	Diseased	NA	China	Asia
SRR5279228	56210438	SRS1997007	SRP100575	Diseased	NA	China	Asia
SRR5279234	52553668	SRS1997013	SRP100575	Diseased	NA	China	Asia
SRR5279236	59536514	SRS1997015	SRP100575	Diseased	NA	China	Asia
SRR5279240	59224720	SRS1997019	SRP100575	Diseased	NA	China	Asia
SRR5279246	50417522	SRS1997025	SRP100575	Diseased	NA	China	Asia
SRR5279252	56987778	SRS1997031	SRP100575	Diseased	NA	China	Asia
SRR5279253	89438324	SRS1997032	SRP100575	Diseased	NA	China	Asia
SRR5279258	80961212	SRS1997037	SRP100575	Diseased	NA	China	Asia
SRR5279267	56193170	SRS1997046	SRP100575	Diseased	NA	China	Asia
SRR5279268	50475514	SRS1997047	SRP100575	Diseased	NA	China	Asia
SRR5279269	87971342	SRS1997048	SRP100575	Diseased	NA	China	Asia
SRR5279271	51433208	SRS1997050	SRP100575	Diseased	NA	China	Asia
SRR5279288	71085312	SRS1997068	SRP100575	Diseased	NA	China	Asia
SRR5279289	51150214	SRS1997067	SRP100575	Diseased	NA	China	Asia
SRR5279290	58554104	SRS1997069	SRP100575	Diseased	NA	China	Asia
SRR5279302	50669070	SRS1997082	SRP100575	Diseased	NA	China	Asia
SRR5963133	60739010	SRS2455728	SRP114966	Diseased	Urban	United States	North America
SRR5963134	57847214	SRS2455729	SRP114966	Diseased	Urban	United States	North America
SRR5963136	66722674	SRS2455731	SRP114966	Diseased	Urban	United States	North America
SRR5963137	56738422	SRS2455732	SRP114966	Diseased	Urban	United States	North America
SRR5963140	65137900	SRS2455735	SRP114966	Diseased	Urban	United States	North America
SRR5963142	50849156	SRS2455737	SRP114966	Diseased	Urban	United States	North America

SRR5963145	54079376	SRS2455740	SRP114966	Diseased	Urban	United States	North America
SRR5963146	59369046	SRS2455741	SRP114966	Diseased	Urban	United States	North America
SRR5963147	94267260	SRS2455742	SRP114966	Diseased	Urban	United States	North America
SRR5963148	71525778	SRS2455743	SRP114966	Diseased	Urban	United States	North America
SRR5963149	51207964	SRS2455744	SRP114966	Diseased	Urban	United States	North America
SRR5963152	55337352	SRS2455747	SRP114966	Diseased	Urban	United States	North America
SRR5963154	58443984	SRS2455749	SRP114966	Diseased	Urban	United States	North America
SRR5963156	51216502	SRS2455752	SRP114966	Diseased	Urban	United States	North America
SRR5963157	56321102	SRS2455751	SRP114966	Diseased	Urban	United States	North America
SRR5963159	63053880	SRS2455754	SRP114966	Diseased	Urban	United States	North America
SRR5963160	50154586	SRS2455755	SRP114966	Diseased	Urban	United States	North America
SRR5963167	51984250	SRS2455762	SRP114966	Diseased	Urban	United States	North America
SRR5963169	50653344	SRS2455764	SRP114966	Diseased	Urban	United States	North America
SRR5963172	161488710	SRS2455767	SRP114966	Diseased	Urban	United States	North America
SRR5963173	93889556	SRS2455768	SRP114966	Diseased	Urban	United States	North America
SRR5963176	89081652	SRS2455771	SRP114966	Diseased	Urban	United States	North America
SRR5963177	57432706	SRS2455772	SRP114966	Diseased	Urban	United States	North America
SRR5963178	56307462	SRS2455773	SRP114966	Diseased	Urban	United States	North America
SRR5963179	57290982	SRS2455774	SRP114966	Diseased	Urban	United States	North America
SRR5963182	53916840	SRS2455777	SRP114966	Diseased	Urban	United States	North America
SRR5963184	70253668	SRS2455779	SRP114966	Diseased	Urban	United States	North America
SRR5963187	51826662	SRS2455782	SRP114966	Diseased	Urban	United States	North America
SRR5963188	54298344	SRS2455783	SRP114966	Diseased	Urban	United States	North America
SRR5963189	50408142	SRS2455784	SRP114966	Diseased	Urban	United States	North America
SRR5963190	53856704	SRS2455785	SRP114966	Diseased	Urban	United States	North America
SRR5963191	60815814	SRS2455787	SRP114966	Diseased	Urban	United States	North America
SRR5963192	55008412	SRS2455788	SRP114966	Diseased	Urban	United States	North America
SRR5963196	50519896	SRS2455791	SRP114966	Diseased	Urban	United States	North America
SRR5963197	54053802	SRS2455792	SRP114966	Diseased	Urban	United States	North America
SRR5963201	167777744	SRS2455797	SRP114966	Diseased	Urban	United States	North America
SRR5963202	92376356	SRS2455796	SRP114966	Diseased	Urban	United States	North America
SRR5963206	53956036	SRS2455801	SRP114966	Diseased	Urban	United States	North America
SRR5963207	177969258	SRS2455802	SRP114966	Diseased	Urban	United States	North America
SRR5963208	106459580	SRS2455803	SRP114966	Diseased	Urban	United States	North America
SRR5963209	53934780	SRS2455804	SRP114966	Diseased	Urban	United States	North America
SRR5963211	56634572	SRS2455806	SRP114966	Diseased	Urban	United States	North America
SRR5963217	50219292	SRS2455812	SRP114966	Diseased	Urban	United States	North America
SRR5963218	56379658	SRS2455813	SRP114966	Diseased	Urban	United States	North America
SRR5963219	67576652	SRS2455815	SRP114966	Diseased	Urban	United States	North America
SRR5963220	69962850	SRS2455814	SRP114966	Diseased	Urban	United States	North America
SRR5963222	55266126	SRS2455817	SRP114966	Diseased	Urban	United States	North America
SRR5963227	52446650	SRS2455822	SRP114966	Diseased	Urban	United States	North America
SRR5963228	51016138	SRS2455823	SRP114966	Diseased	Urban	United States	North America
SRR5963229	52766374	SRS2455825	SRP114966	Diseased	Urban	United States	North America
SRR5963231	53823666	SRS2455826	SRP114966	Diseased	Urban	United States	North America
SRR5963236	53531058	SRS2455830	SRP114966	Diseased	Urban	United States	North America
SRR5963237	51172934	SRS2455832	SRP114966	Diseased	Urban	United States	North America
SRR5963242	50413658	SRS2455838	SRP114966	Diseased	Urban	United States	North America
SRR5963248	55623108	SRS2455843	SRP114966	Diseased	Urban	United States	North America
SRR5963249	53392424	SRS2455844	SRP114966	Diseased	Urban	United States	North America
SRR5963251	80934218	SRS2455847	SRP114966	Diseased	Urban	United States	North America
SRR5963252	57366928	SRS2455845	SRP114966	Diseased	Urban	United States	North America
SRR5963259	50456226	SRS2455855	SRP114966	Diseased	Urban	United States	North America
SRR5963260	51626542	SRS2455854	SRP114966	Diseased	Urban	United States	North America
SRR5963262	52608486	SRS2455857	SRP114966	Diseased	Urban	United States	North America
SRR5963263	51610682	SRS2455858	SRP114966	Diseased	Urban	United States	North America
SRR5963264	56764340	SRS2455859	SRP114966	Diseased	Urban	United States	North America
SRR5963265	55603054	SRS2455860	SRP114966	Diseased	Urban	United States	North America
SRR5963266	56542302	SRS2455861	SRP114966	Diseased	Urban	United States	North America
SRR5963270	56009496	SRS2455865	SRP114966	Diseased	Urban	United States	North America
SRR5963271	59830092	SRS2455866	SRP114966	Diseased	Urban	United States	North America
SRR5963272	57254692	SRS2455867	SRP114966	Diseased	Urban	United States	North America
SRR5963277	52758498	SRS2455873	SRP114966	Diseased	Urban	United States	North America
SRR5963279	51309362	SRS2455874	SRP114966	Diseased	Urban	United States	North America
SRR5963280	78760854	SRS2455875	SRP114966	Diseased	Urban	United States	North America
SRR5963281	53692284	SRS2455876	SRP114966	Diseased	Urban	United States	North America
SRR5963282	50071006	SRS2455877	SRP114966	Diseased	Urban	United States	North America
SRR5963283	55268596	SRS2455878	SRP114966	Diseased	Urban	United States	North America
SRR5963284	50341042	SRS2455879	SRP114966	Diseased	Urban	United States	North America
SRR5963297	60155628	SRS2455892	SRP114966	Diseased	Urban	United States	North America
SRR5963298	56558116	SRS2455893	SRP114966	Diseased	Urban	United States	North America
SRR5963299	55502360	SRS2455894	SRP114966	Diseased	Urban	United States	North America

SRR5963300	64883388	SRS2455896	SRP114966	Diseased	Urban	United States	North America
SRR5963302	61758138	SRS2455897	SRP114966	Diseased	Urban	United States	North America
SRR5963305	55859292	SRS2455900	SRP114966	Diseased	Urban	United States	North America
SRR5963307	59862060	SRS2455902	SRP114966	Diseased	Urban	United States	North America
SRR5963316	61440438	SRS2455911	SRP114966	Diseased	Urban	United States	North America
SRR5963318	52692600	SRS2455913	SRP114966	Diseased	Urban	United States	North America
SRR5963319	54722204	SRS2455914	SRP114966	Diseased	Urban	United States	North America
SRR5963320	50847606	SRS2455917	SRP114966	Diseased	Urban	United States	North America
SRR5963328	52724468	SRS2455923	SRP114966	Diseased	Urban	United States	North America
SRR5963329	53857528	SRS2455924	SRP114966	Diseased	Urban	United States	North America
SRR5963330	53183354	SRS2455925	SRP114966	Diseased	Urban	United States	North America
SRR5963355	55774846	SRS2455951	SRP114966	Diseased	Urban	United States	North America
SRR5963358	51241404	SRS2455953	SRP114966	Diseased	Urban	United States	North America
SRR5963360	50250108	SRS2455955	SRP114966	Diseased	Urban	United States	North America
SRR5963365	56262004	SRS2455960	SRP114966	Diseased	Urban	United States	North America
SRR5963368	61501414	SRS2455963	SRP114966	Diseased	Urban	United States	North America
SRR5963369	64541284	SRS2455965	SRP114966	Diseased	Urban	United States	North America
SRR5963371	51909100	SRS2455967	SRP114966	Diseased	Urban	United States	North America
SRR5963372	65850174	SRS2455968	SRP114966	Diseased	Urban	United States	North America
SRR5963373	58884958	SRS2455966	SRP114966	Diseased	Urban	United States	North America
SRR5963376	50412440	SRS2455971	SRP114966	Diseased	Urban	United States	North America
SRR5963377	52316486	SRS2455972	SRP114966	Diseased	Urban	United States	North America
SRR5963378	84922508	SRS2455973	SRP114966	Diseased	Urban	United States	North America
SRR5963379	156682726	SRS2455974	SRP114966	Diseased	Urban	United States	North America
SRR5963380	285255994	SRS2455975	SRP114966	Diseased	Urban	United States	North America
SRR5963381	145466986	SRS2455976	SRP114966	Diseased	Urban	United States	North America
SRR5963382	149744824	SRS2455977	SRP114966	Diseased	Urban	United States	North America
SRR5963383	363480576	SRS2455978	SRP114966	Diseased	Urban	United States	North America
SRR5963384	274156614	SRS2455979	SRP114966	Diseased	Urban	United States	North America
SRR5963385	208389686	SRS2455980	SRP114966	Diseased	Urban	United States	North America
SRR5963386	286050396	SRS2455981	SRP114966	Diseased	Urban	United States	North America
SRR5963387	390694326	SRS2455982	SRP114966	Diseased	Urban	United States	North America
SRR5963388	352745808	SRS2455983	SRP114966	Diseased	Urban	United States	North America
SRR5963389	59226246	SRS2455984	SRP114966	Diseased	Urban	United States	North America
SRR5963390	199401778	SRS2455985	SRP114966	Diseased	Urban	United States	North America
SRR5963391	206457966	SRS2455986	SRP114966	Diseased	Urban	United States	North America
SRR5963393	217298342	SRS2455988	SRP114966	Diseased	Urban	United States	North America
SRR5963394	442429912	SRS2455989	SRP114966	Diseased	Urban	United States	North America
SRR5963395	540444744	SRS2455991	SRP114966	Diseased	Urban	United States	North America
SRR5963396	215119022	SRS2455990	SRP114966	Diseased	Urban	United States	North America
SRR5963397	436079720	SRS2455992	SRP114966	Diseased	Urban	United States	North America
SRR5963398	202455212	SRS2455993	SRP114966	Diseased	Urban	United States	North America
SRR5963399	262465788	SRS2455994	SRP114966	Diseased	Urban	United States	North America
SRR5963400	289947154	SRS2455996	SRP114966	Diseased	Urban	United States	North America
SRR5963401	134350242	SRS2455995	SRP114966	Diseased	Urban	United States	North America
SRR5963402	190025612	SRS2455997	SRP114966	Diseased	Urban	United States	North America
SRR5963403	148354464	SRS2456000	SRP114966	Diseased	Urban	United States	North America
SRR5963404	234045426	SRS2455999	SRP114966	Diseased	Urban	United States	North America
SRR5963405	130742362	SRS2456001	SRP114966	Diseased	Urban	United States	North America
SRR5963406	92603342	SRS2455998	SRP114966	Diseased	Urban	United States	North America
SRR5963407	204321348	SRS2456002	SRP114966	Diseased	Urban	United States	North America
SRR5963408	229828250	SRS2456003	SRP114966	Diseased	Urban	United States	North America
SRR5963409	123280528	SRS2456004	SRP114966	Diseased	Urban	United States	North America
SRR5963410	220075262	SRS2456005	SRP114966	Diseased	Urban	United States	North America
SRR5963411	140340764	SRS2456006	SRP114966	Diseased	Urban	United States	North America
SRR5963412	217181326	SRS2456009	SRP114966	Diseased	Urban	United States	North America
SRR5963413	209372928	SRS2456008	SRP114966	Diseased	Urban	United States	North America
SRR5963414	175552856	SRS2456007	SRP114966	Diseased	Urban	United States	North America
SRR5963415	136960042	SRS2456010	SRP114966	Diseased	Urban	United States	North America
SRR5963417	172216104	SRS2456012	SRP114966	Diseased	Urban	United States	North America
SRR5963418	230361800	SRS2456013	SRP114966	Diseased	Urban	United States	North America
SRR5963419	136148810	SRS2456014	SRP114966	Diseased	Urban	United States	North America
SRR5963420	223052674	SRS2456015	SRP114966	Diseased	Urban	United States	North America
SRR5963422	245197202	SRS2456017	SRP114966	Diseased	Urban	United States	North America
SRR5963423	203228158	SRS2456018	SRP114966	Diseased	Urban	United States	North America
SRR5963424	189533812	SRS2456019	SRP114966	Diseased	Urban	United States	North America
SRR5963425	189365090	SRS2456020	SRP114966	Diseased	Urban	United States	North America
SRR5963426	186895654	SRS2456023	SRP114966	Diseased	Urban	United States	North America
SRR5963427	164361808	SRS2456021	SRP114966	Diseased	Urban	United States	North America
SRR5963428	287792224	SRS2456022	SRP114966	Diseased	Urban	United States	North America
SRR5963429	257460938	SRS2456024	SRP114966	Diseased	Urban	United States	North America
SRR5963430	85405310	SRS2456025	SRP114966	Diseased	Urban	United States	North America

SRR5963431	133703426	SRS2456026	SRP114966	Diseased	Urban	United States	North America
SRR5963432	270284910	SRS2456027	SRP114966	Diseased	Urban	United States	North America
SRR5963433	313678716	SRS2456028	SRP114966	Diseased	Urban	United States	North America
SRR5963434	143271554	SRS2456029	SRP114966	Diseased	Urban	United States	North America
SRR7658579	76164016	SRS3638681	SRP156699	Healthy	Rural	Madagascar	Africa
SRR7658580	78891192	SRS3638680	SRP156699	Healthy	Rural	Madagascar	Africa
SRR7658581	72309958	SRS3638679	SRP156699	Healthy	Rural	Madagascar	Africa
SRR7658582	76314684	SRS3638678	SRP156699	Healthy	Rural	Madagascar	Africa
SRR7658583	76639120	SRS3638676	SRP156699	Healthy	Rural	Madagascar	Africa
SRR7658584	76552676	SRS3638677	SRP156699	Healthy	Rural	Madagascar	Africa
SRR7658585	68096090	SRS3638675	SRP156699	Healthy	Rural	Madagascar	Africa
SRR7658586	67713040	SRS3638673	SRP156699	Healthy	Rural	Madagascar	Africa
SRR7658587	69793746	SRS3638672	SRP156699	Healthy	Rural	Madagascar	Africa
SRR7658588	70322274	SRS3638671	SRP156699	Healthy	Rural	Madagascar	Africa
SRR7658627	50601868	SRS3638628	SRP156699	Healthy	Rural	Madagascar	Africa
SRR7658639	122581570	SRS3638617	SRP156699	Healthy	Rural	Madagascar	Africa
SRR7658640	145454712	SRS3638614	SRP156699	Healthy	Rural	Madagascar	Africa
SRR7658641	108295878	SRS3638615	SRP156699	Healthy	Rural	Madagascar	Africa
SRR7658642	110337100	SRS3638613	SRP156699	Healthy	Rural	Madagascar	Africa
SRR7658643	105329262	SRS3638612	SRP156699	Healthy	Rural	Madagascar	Africa
SRR7658644	106117224	SRS3638611	SRP156699	Healthy	Rural	Madagascar	Africa
SRR7658645	115980674	SRS3638610	SRP156699	Healthy	Rural	Madagascar	Africa
SRR7658646	122563118	SRS3638609	SRP156699	Healthy	Rural	Madagascar	Africa
SRR7658647	116564528	SRS3638608	SRP156699	Healthy	Rural	Madagascar	Africa
SRR7658648	116435670	SRS3638607	SRP156699	Healthy	Rural	Madagascar	Africa
SRR7658649	67399482	SRS3638606	SRP156699	Healthy	Rural	Madagascar	Africa
SRR7658650	67182444	SRS3638605	SRP156699	Healthy	Rural	Madagascar	Africa
SRR7658651	66307656	SRS3638604	SRP156699	Healthy	Rural	Madagascar	Africa
SRR7658652	65227972	SRS3638682	SRP156699	Healthy	Rural	Madagascar	Africa
SRR7658653	66839968	SRS3638603	SRP156699	Healthy	Rural	Madagascar	Africa
SRR7658654	66579808	SRS3638601	SRP156699	Healthy	Rural	Madagascar	Africa
SRR7658655	58731516	SRS3638602	SRP156699	Healthy	Rural	Madagascar	Africa
SRR7658656	57287200	SRS3638598	SRP156699	Healthy	Rural	Madagascar	Africa
SRR7658657	60558908	SRS3638600	SRP156699	Healthy	Rural	Madagascar	Africa
SRR7658658	60100950	SRS3638599	SRP156699	Healthy	Rural	Madagascar	Africa
SRR7658659	85522338	SRS3638597	SRP156699	Healthy	Rural	Madagascar	Africa
SRR7658660	83149084	SRS3638595	SRP156699	Healthy	Rural	Madagascar	Africa
SRR7658661	82766976	SRS3638596	SRP156699	Healthy	Rural	Madagascar	Africa
SRR7658662	81972528	SRS3638664	SRP156699	Healthy	Rural	Madagascar	Africa
SRR7658663	90256322	SRS3638594	SRP156699	Healthy	Rural	Madagascar	Africa
SRR7658664	89049582	SRS3638592	SRP156699	Healthy	Rural	Madagascar	Africa
SRR7658665	86625656	SRS3638593	SRP156699	Healthy	Rural	Madagascar	Africa
SRR7658666	85588874	SRS3638591	SRP156699	Healthy	Rural	Madagascar	Africa
SRR7658667	218939862	SRS3638590	SRP156699	Healthy	Rural	Madagascar	Africa
SRR7658668	164844578	SRS3638653	SRP156699	Healthy	Rural	Madagascar	Africa
SRR7658669	97679014	SRS3638589	SRP156699	Healthy	Rural	Madagascar	Africa
SRR7658670	92960178	SRS3638588	SRP156699	Healthy	Rural	Madagascar	Africa
SRR7658671	57165974	SRS3638642	SRP156699	Healthy	Rural	Madagascar	Africa
SRR7658672	57471574	SRS3638587	SRP156699	Healthy	Rural	Madagascar	Africa
SRR7658673	56318136	SRS3638586	SRP156699	Healthy	Rural	Madagascar	Africa
SRR7658674	57002392	SRS3638638	SRP156699	Healthy	Rural	Madagascar	Africa
SRR7658675	53675748	SRS3638585	SRP156699	Healthy	Rural	Madagascar	Africa
SRR7658676	54390734	SRS3638583	SRP156699	Healthy	Rural	Madagascar	Africa
SRR7658677	52133298	SRS3638584	SRP156699	Healthy	Rural	Madagascar	Africa
SRR7658678	52421660	SRS3638582	SRP156699	Healthy	Rural	Madagascar	Africa
SRR7658679	51327308	SRS3638581	SRP156699	Healthy	Rural	Madagascar	Africa
SRR7658680	51850178	SRS3638632	SRP156699	Healthy	Rural	Madagascar	Africa
SRR8180446	70476226	SRS4035281	SRP168387	NA	NA	Ethiopia	Africa
SRR8180447	60270408	SRS4035280	SRP168387	NA	NA	Ethiopia	Africa
SRR8180448	78527636	SRS4035279	SRP168387	NA	NA	Ethiopia	Africa
SRR8180449	72761218	SRS4035278	SRP168387	NA	NA	Ethiopia	Africa
SRR8180450	89872686	SRS4035277	SRP168387	NA	NA	Ethiopia	Africa