



US 20150284814A1

(19) **United States**

(12) **Patent Application Publication**  
**Barouch et al.**

(10) **Pub. No.: US 2015/0284814 A1**  
(43) **Pub. Date: Oct. 8, 2015**

(54) **VIRUSES ASSOCIATED WITH IMMUNODEFICIENCY AND ENTEROPATHY AND METHODS USING SAME**

(86) PCT No.: **PCT/US13/60579**  
§ 371 (c)(1),  
(2) Date: **Mar. 19, 2015**

(71) Applicants: **BETH ISRAEL DEACONESS MEDICAL CENTER, INC.**, Boston, MA (US); **WASHINGTON UNIVERSITY**, Saint Louis, MO (US)

**Related U.S. Application Data**

(60) Provisional application No. 61/702,952, filed on Sep. 19, 2012.

(72) Inventors: **Dan Barouch**, Newton, MA (US); **Herbert W. Virgin**, Saint Louis, MO (US); **David Wang**, Saint Louis, MO (US); **Guoyan Zhao**, Saint Louis, MO (US); **Larissa Thackray**, Columbia, IL (US); **Scott Handley**, Saint Louis, MO (US); **Rachel Presti**, University City, MO (US)

**Publication Classification**

(51) **Int. Cl.**  
**C12Q 1/70** (2006.01)  
(52) **U.S. Cl.**  
CPC ..... **C12Q 1/703** (2013.01); **C12Q 2600/112** (2013.01); **C12Q 2600/158** (2013.01)

(73) Assignee: **Beth Israel Deaconess**, Boston, MA (US)

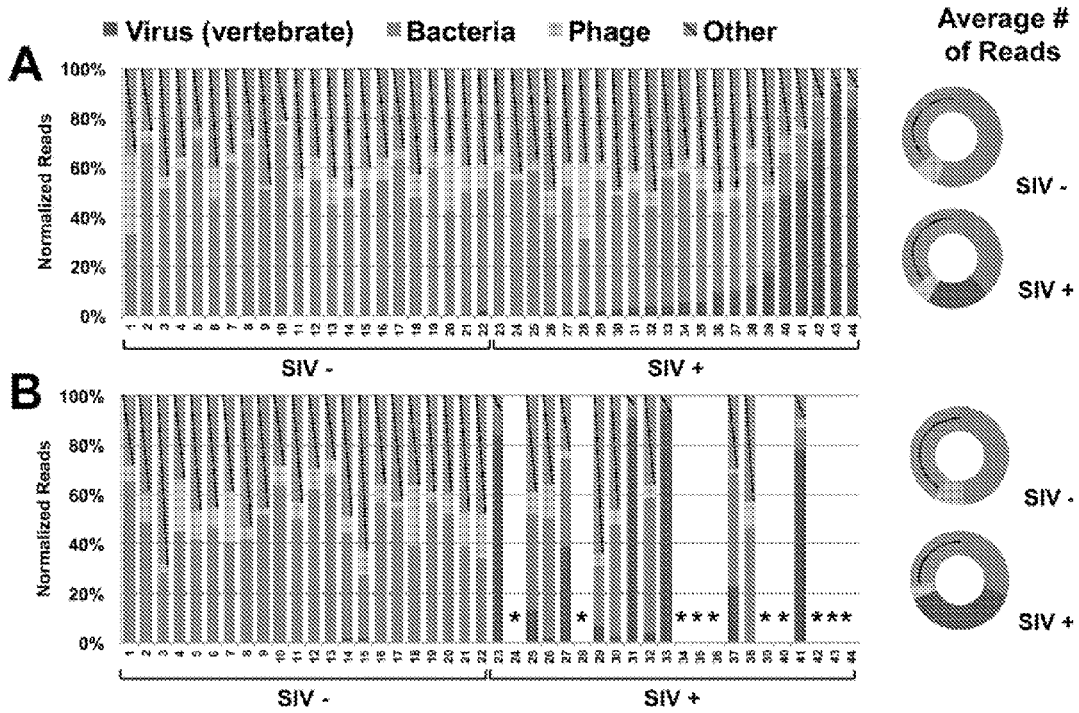
(57) **ABSTRACT**

The present invention relates to previously undescribed viruses that are associated with significant expansion of the virome, immunodeficiency, and enteropathy during lentiviral infection. The invention also provides methods to detect acquired immune deficiency syndrome (AIDS) or AIDS progression in a subject, methods to diagnose immunodeficiency or enteropathy in a subject, and methods to identify a therapeutic agent to treat the same.

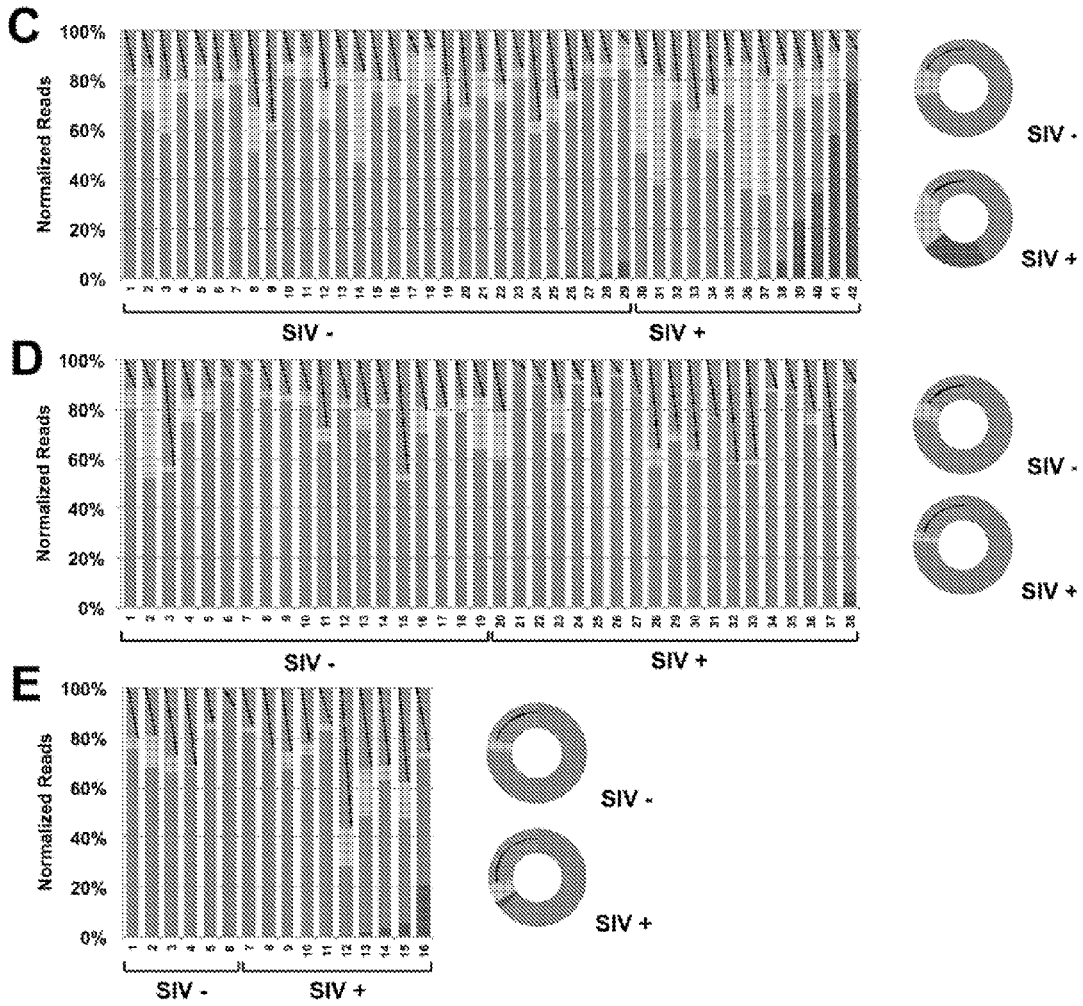
(21) Appl. No.: **14/429,559**

(22) PCT Filed: **Sep. 19, 2013**

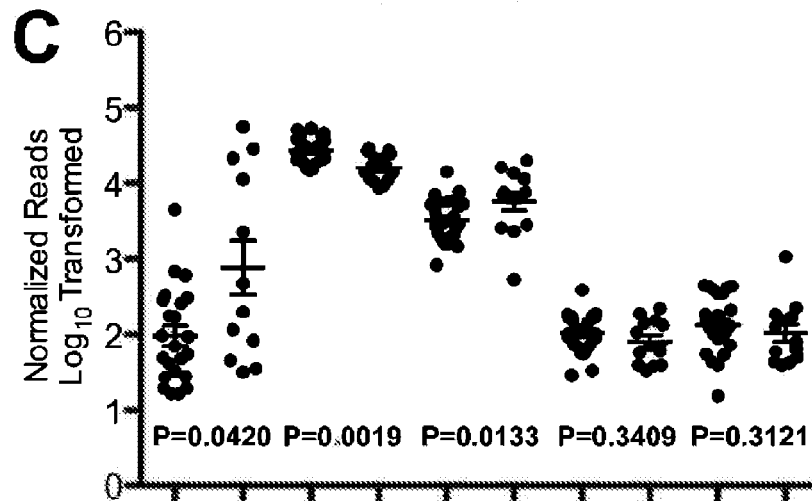
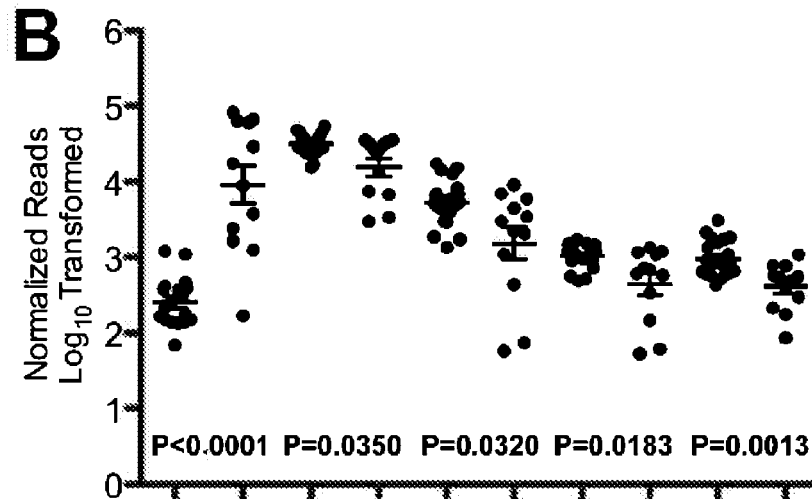
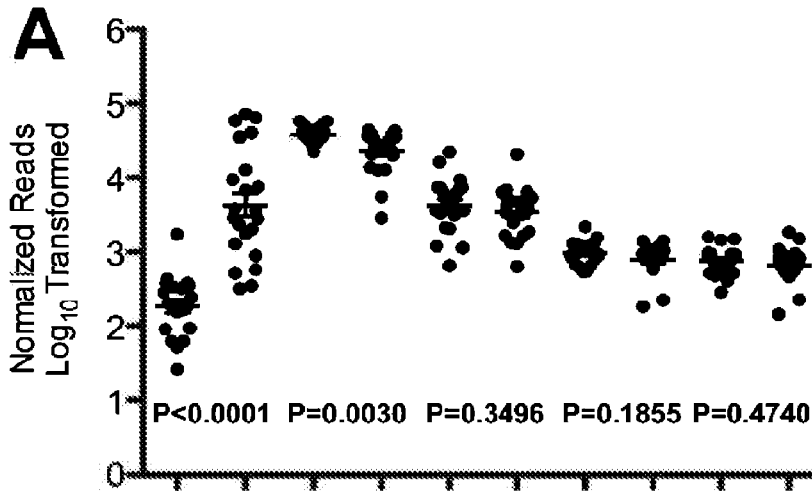
Figures 1A-1B



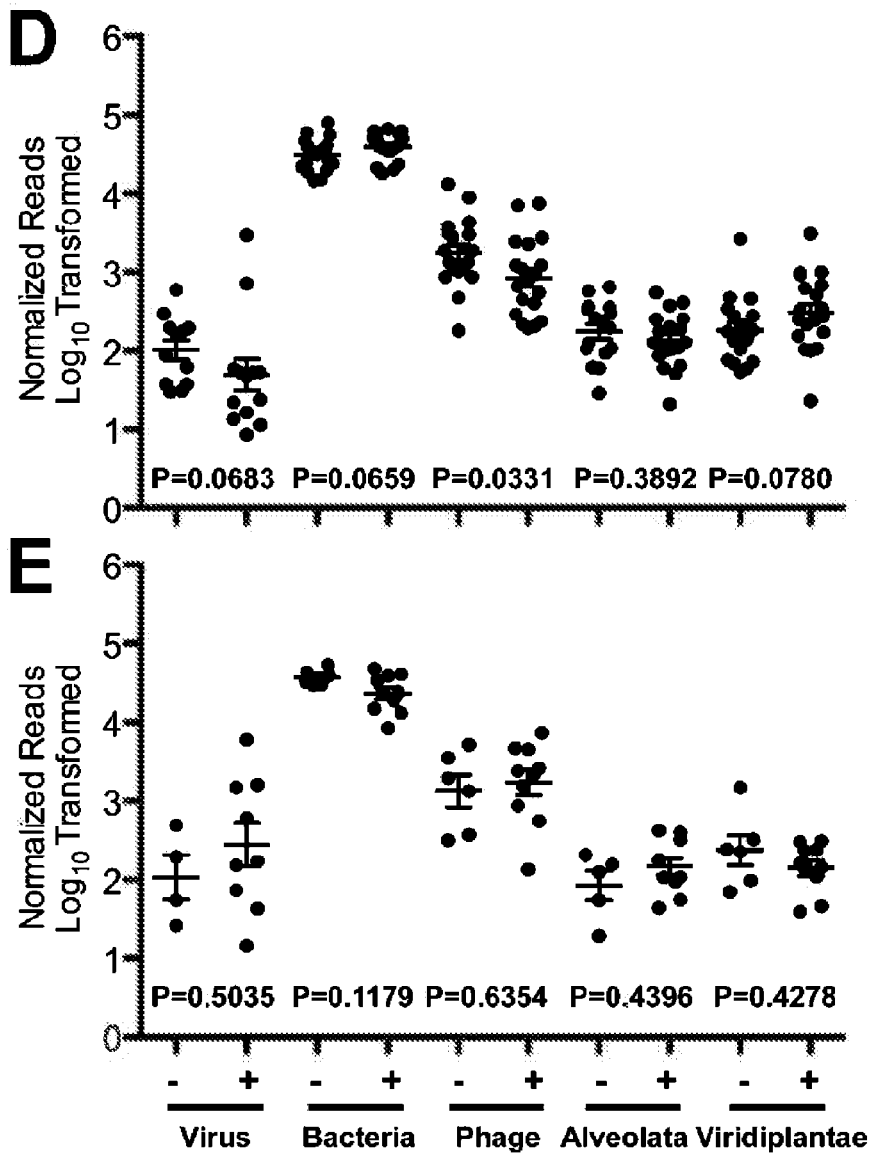
Figures 1C-1E



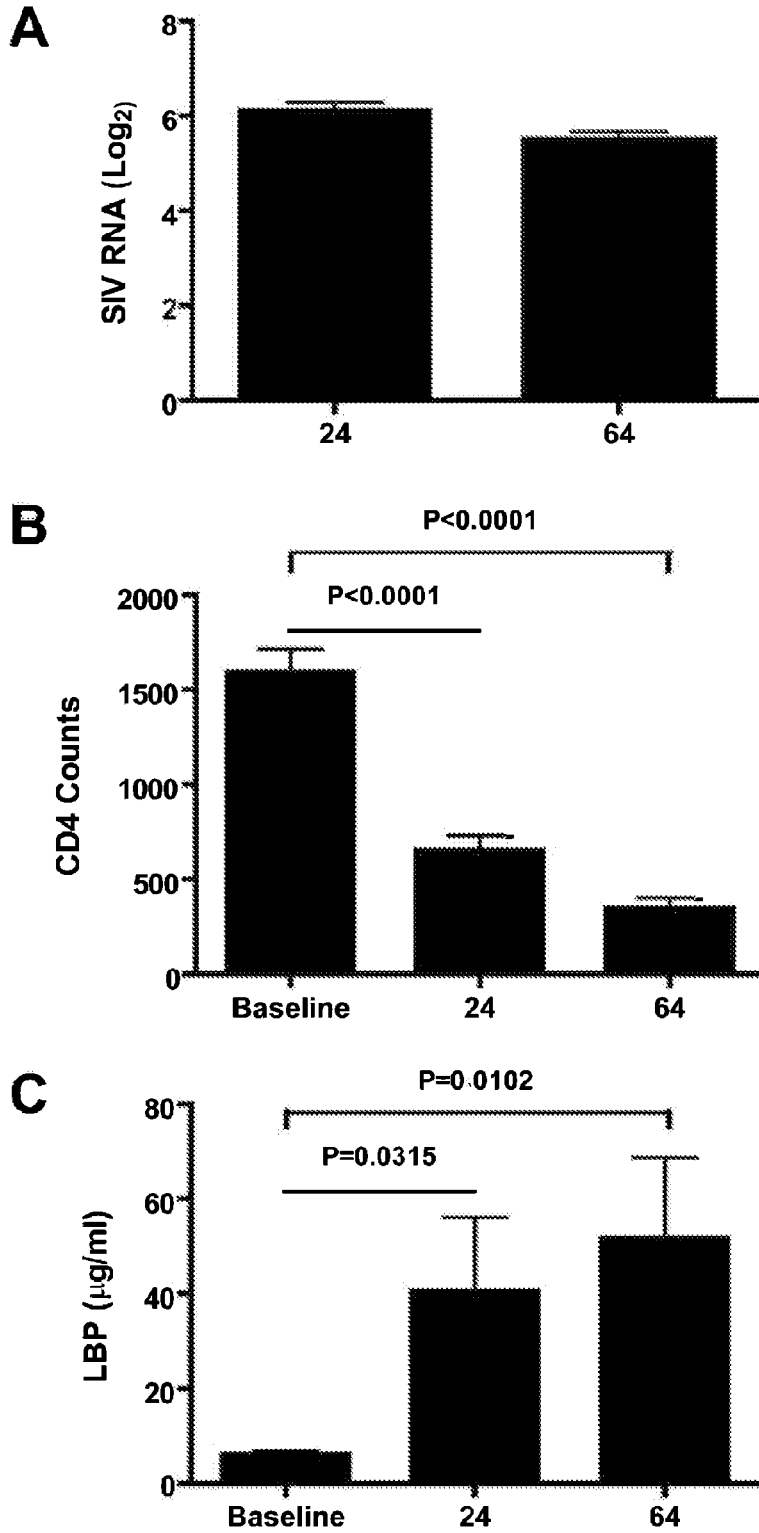
Figures 2A-2C



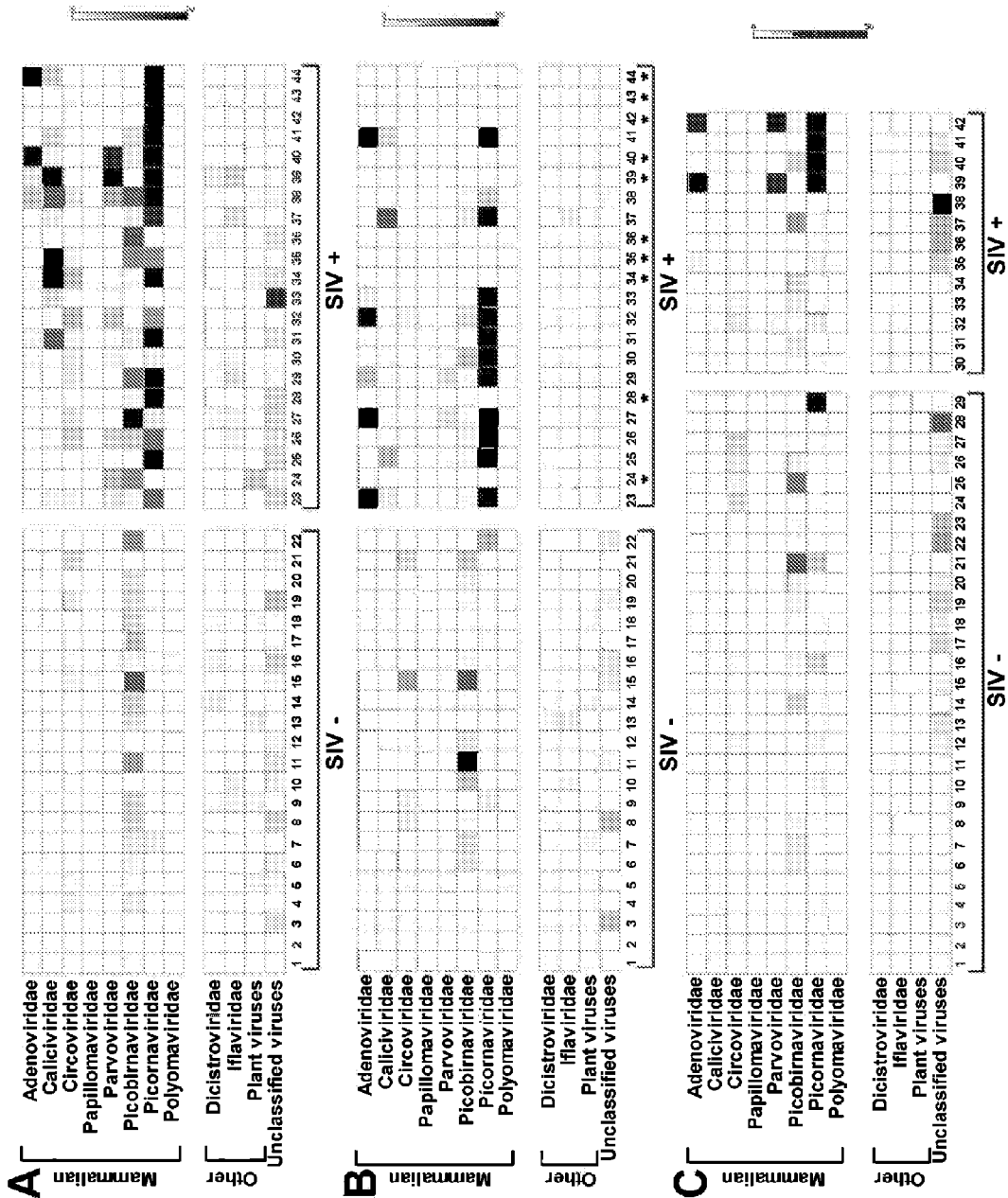
Figures 2D-2E



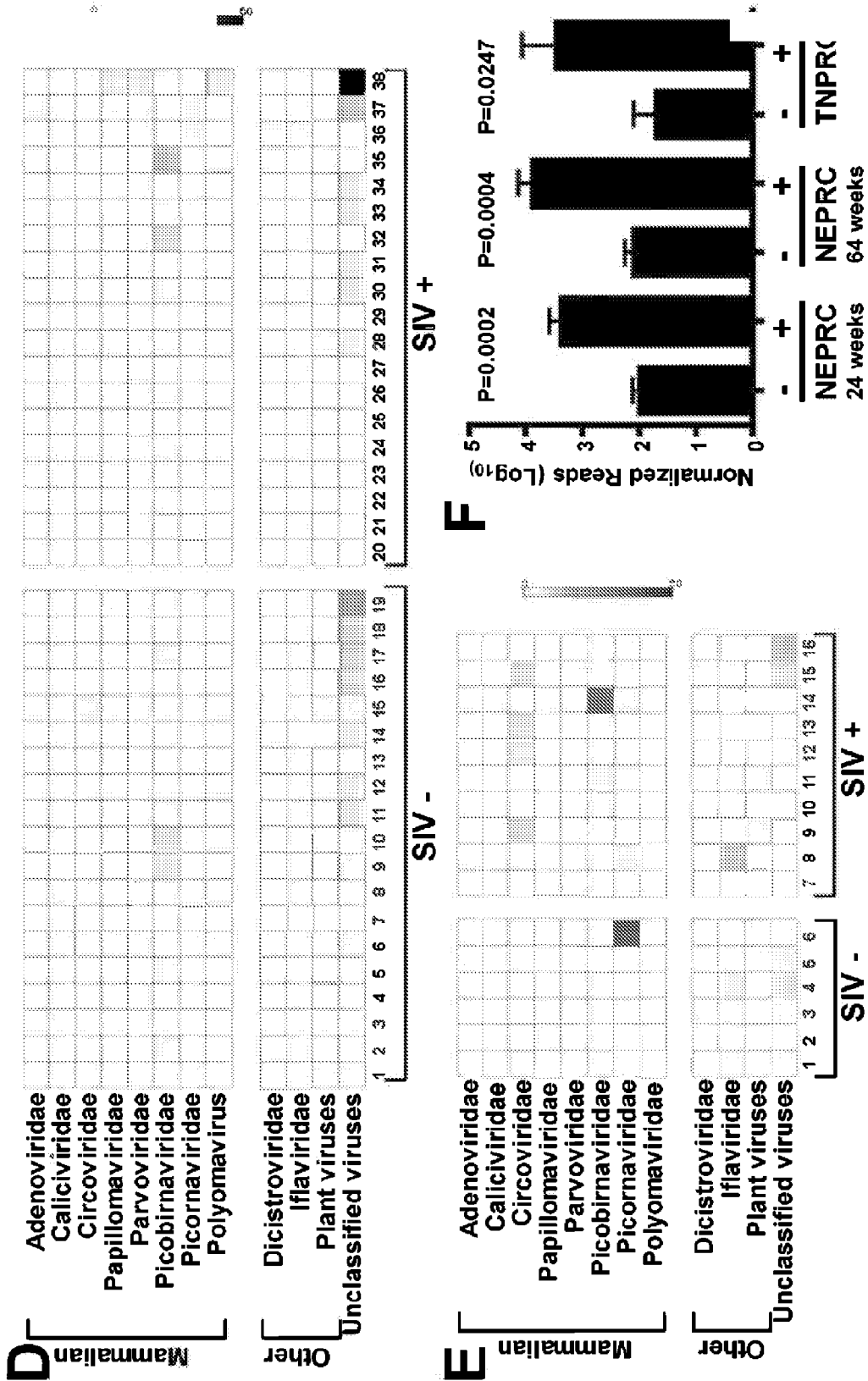
Figures 3A-3C



Figures 4A-4C

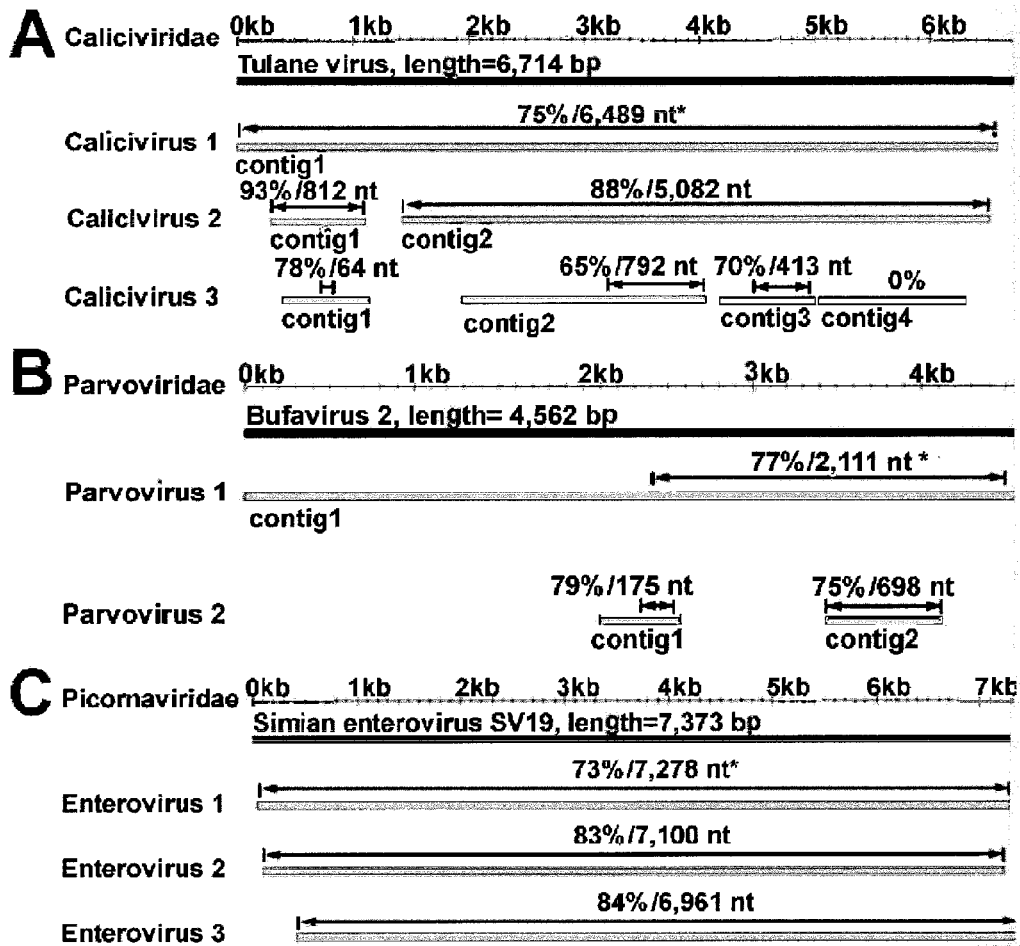


Figures 4D-4F

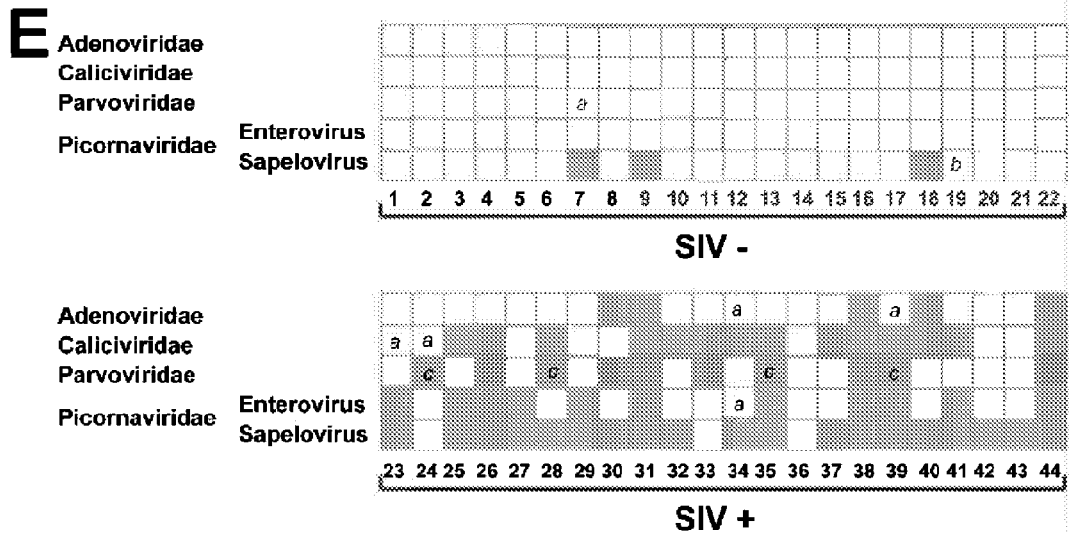
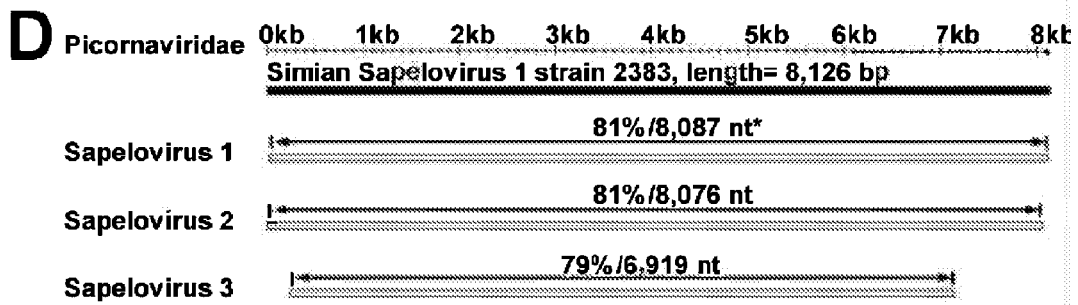




Figures 5A-5C



Figures 5D-5E



Figures 6A-6C

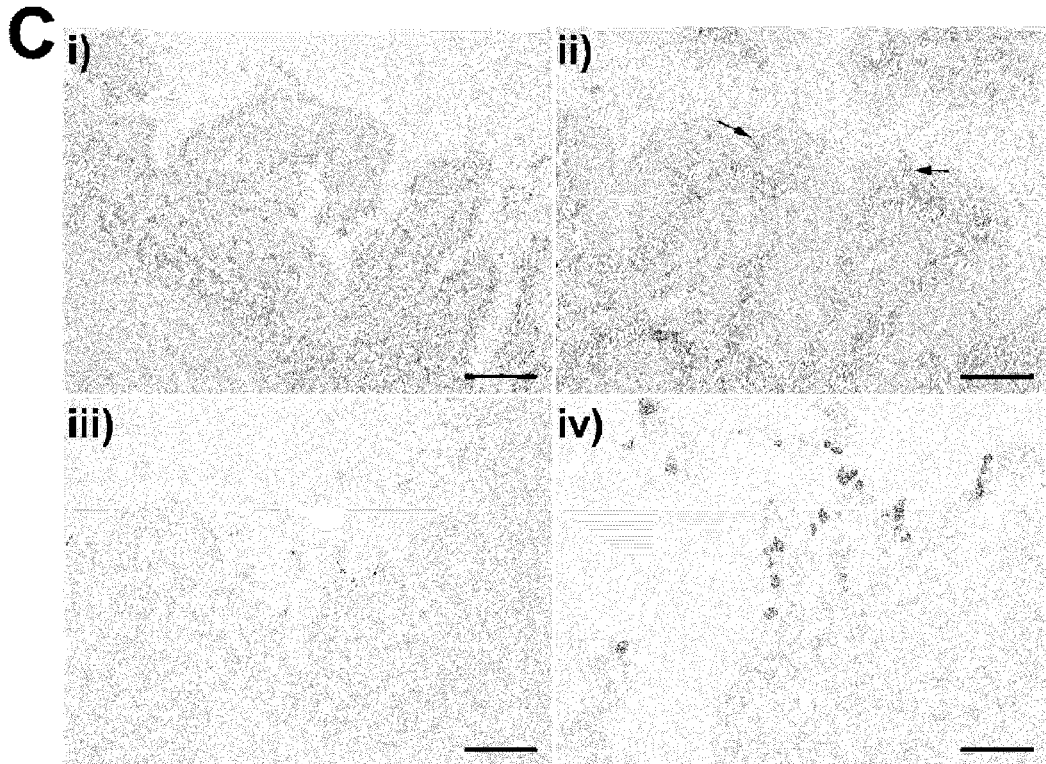
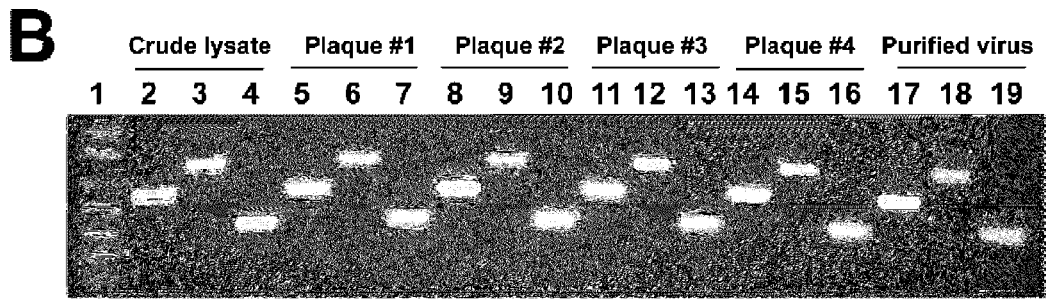
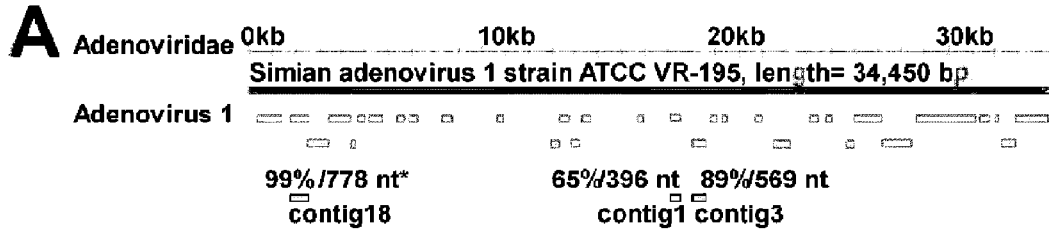
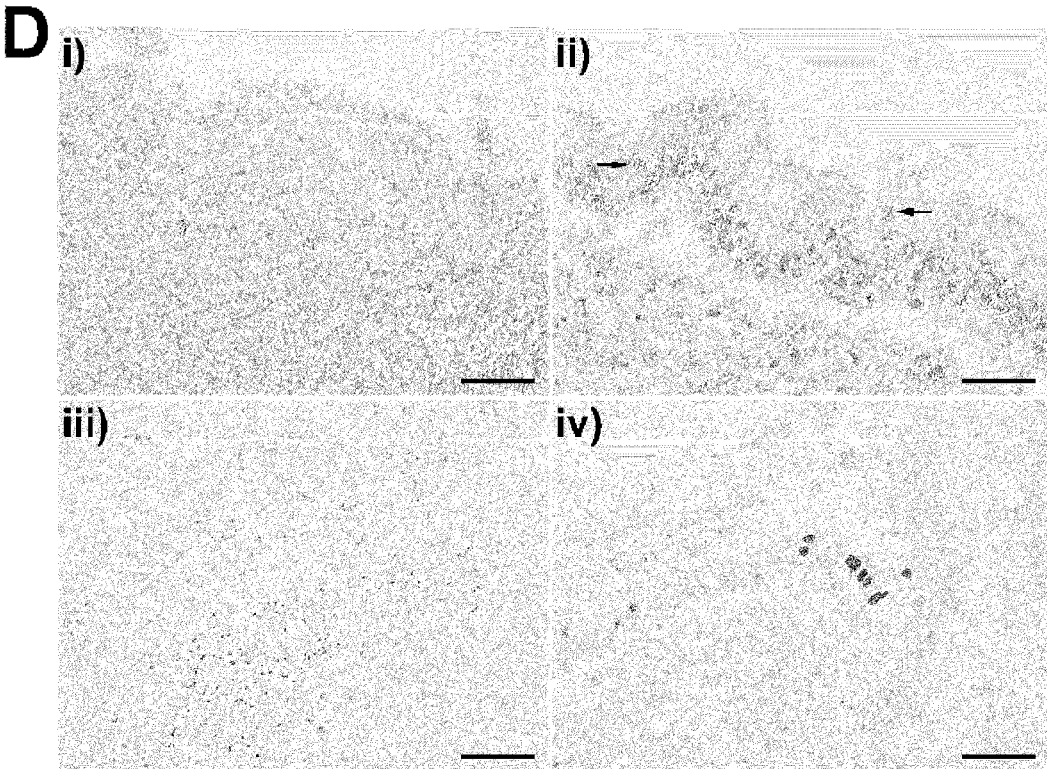
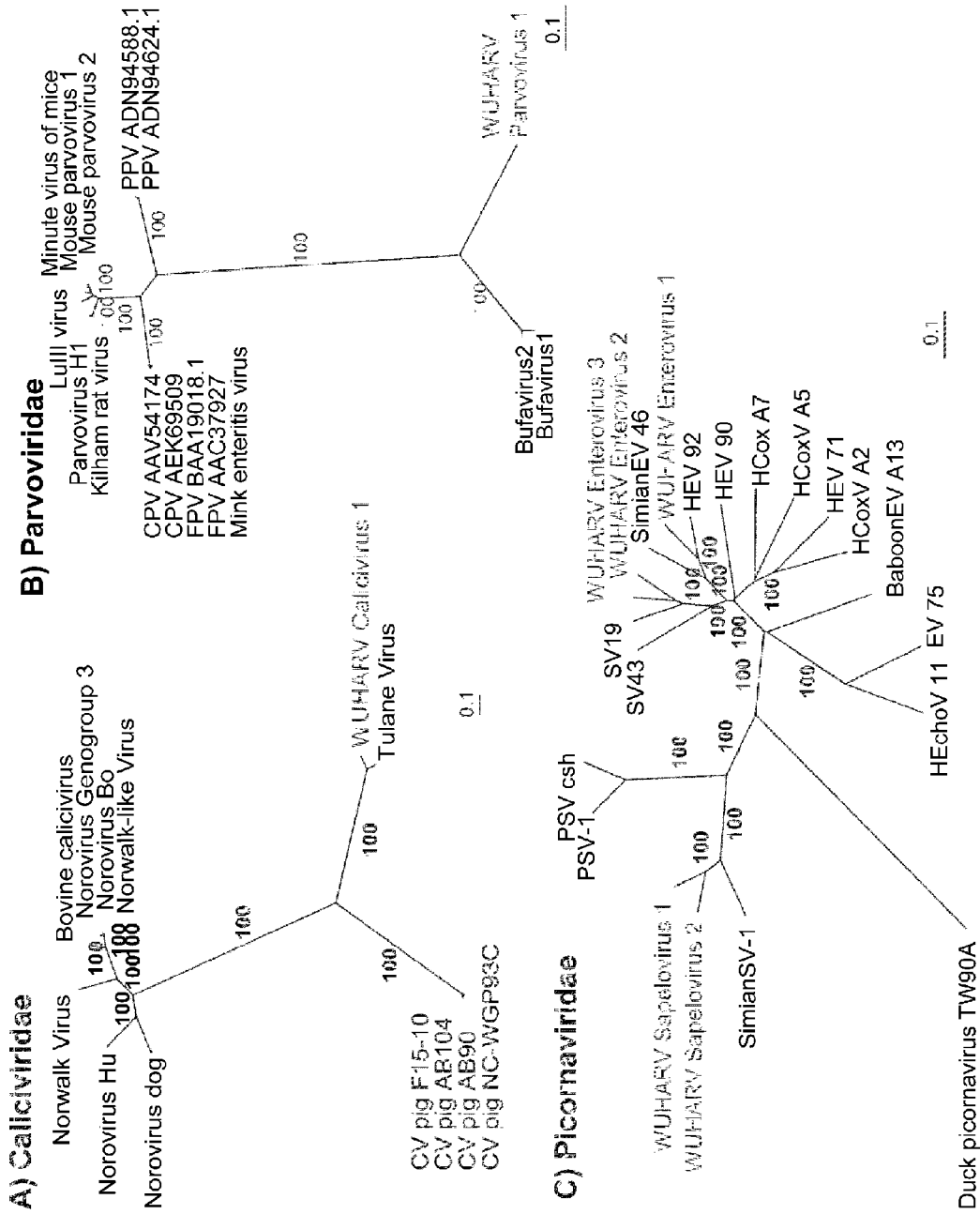


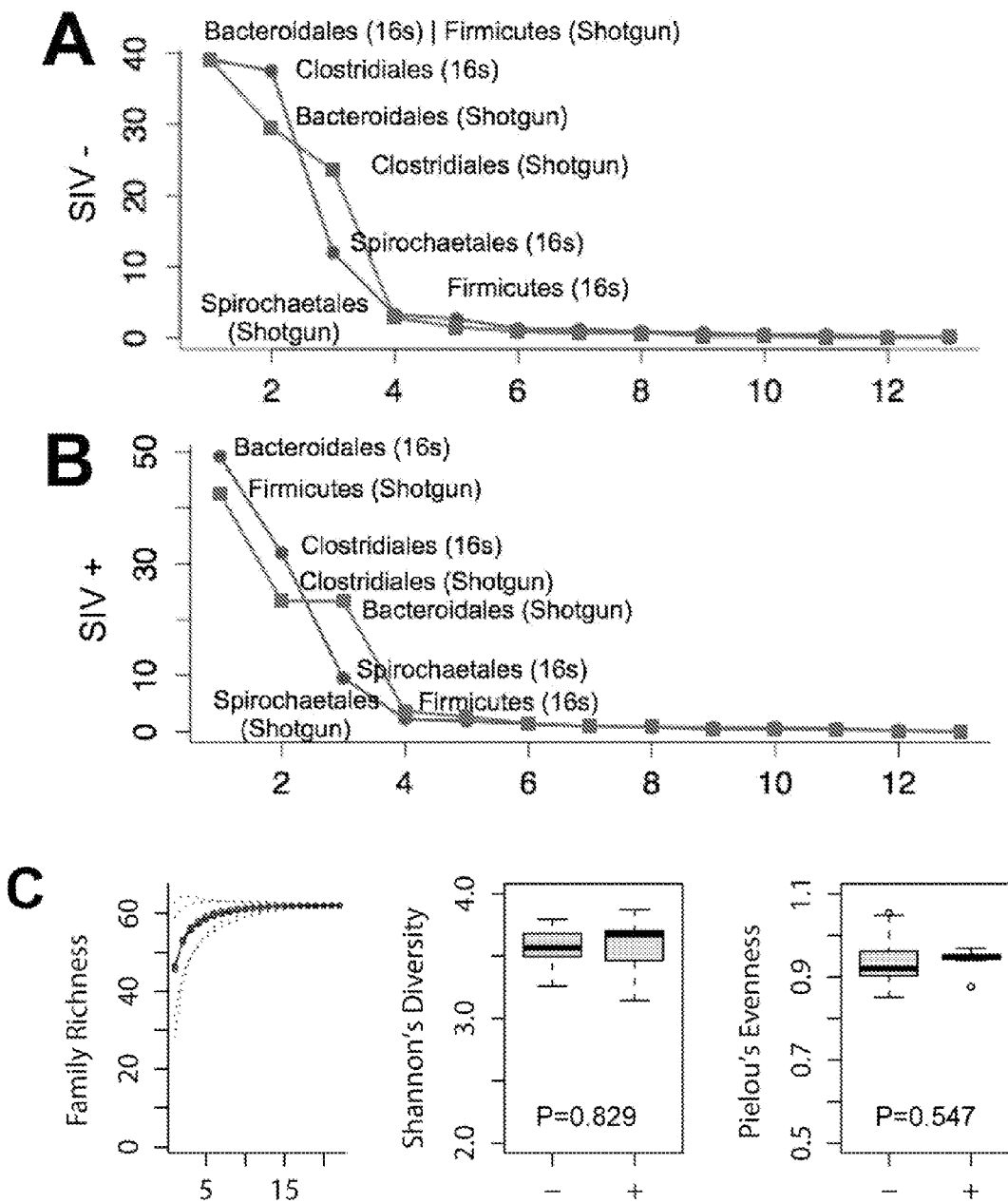
Figure 6D



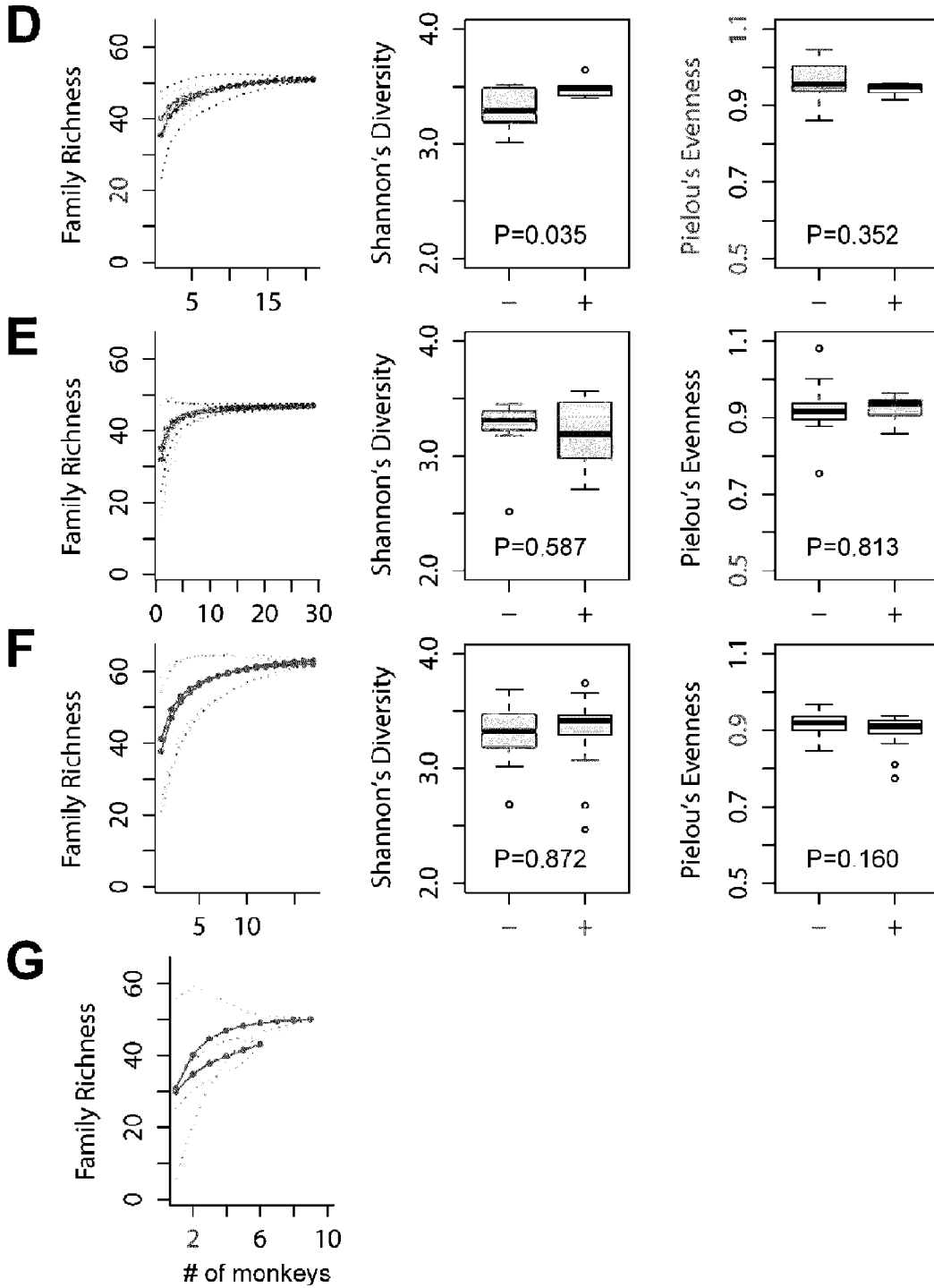
Figures 7A-7C



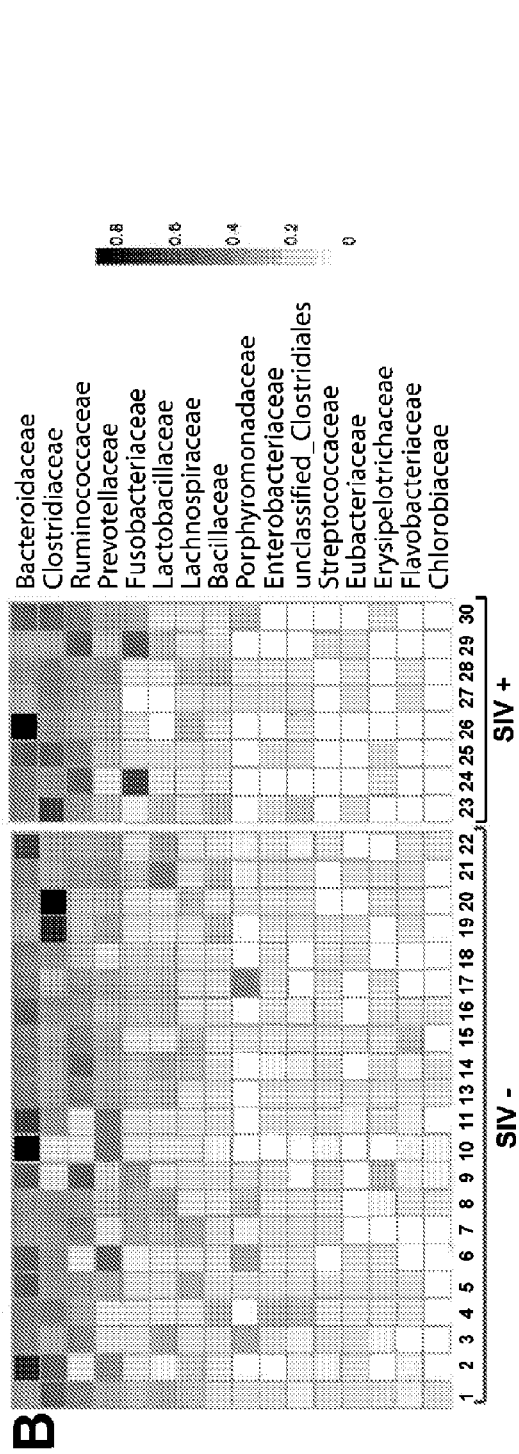
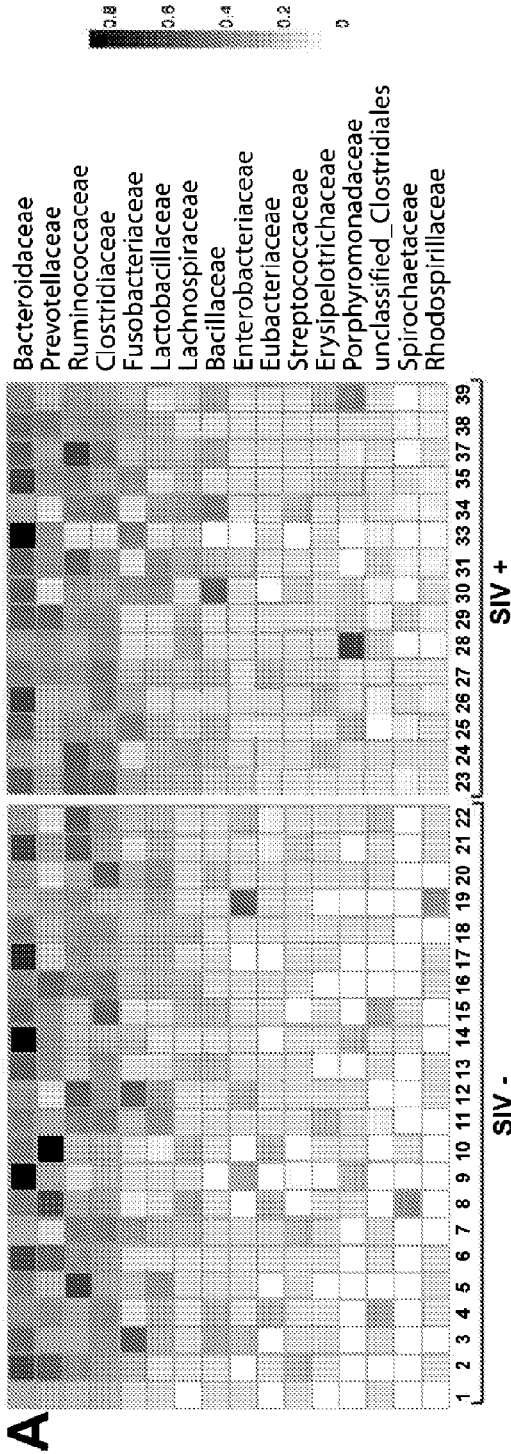
Figures 8A-8C



Figures 8D-8G

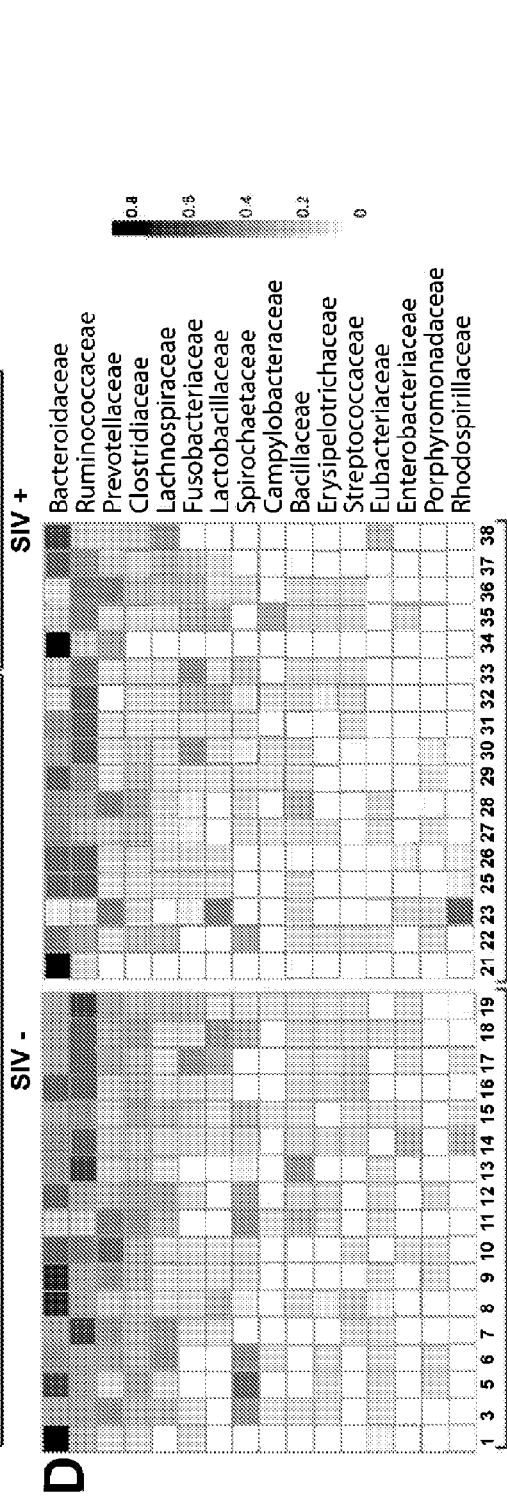
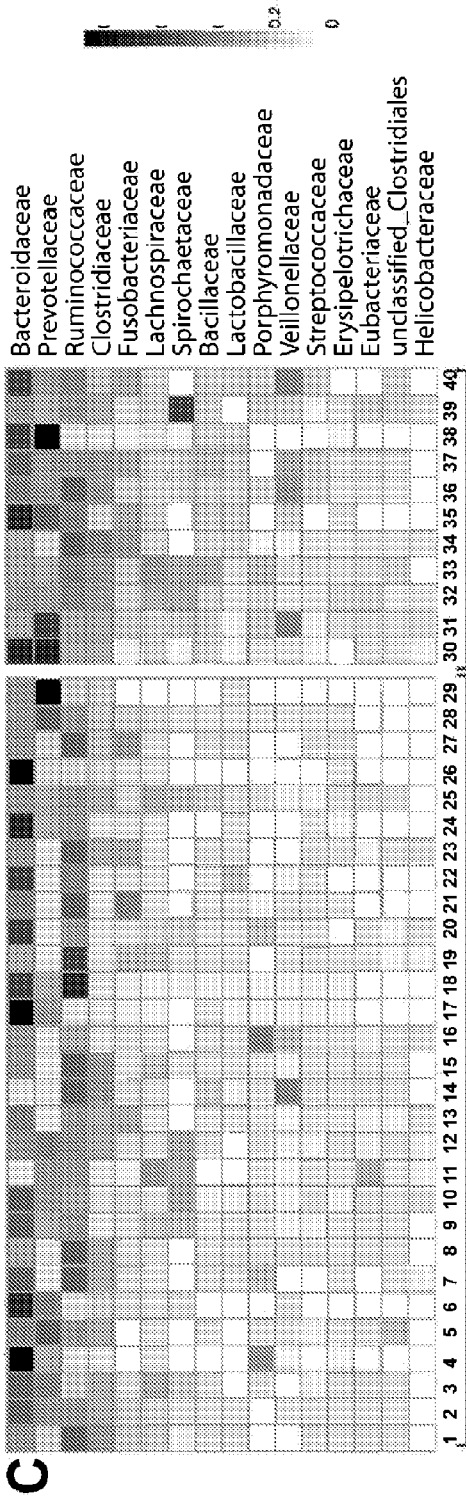


Figures 9A-9B





Figures 9C-9D



## VIRUSES ASSOCIATED WITH IMMUNODEFICIENCY AND ENTEROPATHY AND METHODS USING SAME

### BACKGROUND OF THE INVENTION

**[0001]** Human immunodeficiency virus (HIV) infection of humans and pathogenic simian immunodeficiency virus (SIV) infection of rhesus monkeys causes progressive immunocompromise and acquired immune deficiency syndrome (AIDS). One hallmark that correlates with the rate of progression to AIDS is systemic immune activation. Systemic immune activation is, in turn, associated with damage to the intestinal epithelium (enteropathy) and translocation of as-yet-undefined immunostimulatory pathogen-associated molecular patterns (PAMPS) or antigens into tissues and the blood.

**[0002]** Despite the importance of intestinal barrier damage to AIDS progression, the mechanisms responsible for AIDS enteropathy are not understood. One possibility is that immunodeficiency leads to epithelial damage by intestinal viruses or other pathogens. The mammalian virome and bacterial microbiome is extremely complex and can contribute to immune status and disease in a range of settings. Thus far, a prior study that utilized 16S rDNA sequencing, which was unable to detect viruses, found no discernible differences in the diversity of bacteria associated with SIV infection (McKenna et al., *PLoS Pathog.* 4: e20 (2008)). However, it remained a possibility that the virome, a subset of the metagenome that may be defined as viruses that infect eukaryotic cells, contributes to epithelial damage during lentiviral infection.

**[0003]** It is therefore important to understand the contribution of the virome to lentiviral infection-associated phenotypes, such as enteropathy. There is an unmet need in the field for understanding the contribution of the virome upon lentiviral infection, as well as for the development of alternative methods of diagnosing and treating lentiviral infections (e.g., HIV).

### SUMMARY OF THE INVENTION

**[0004]** This invention relates to the discovery of previously undescribed viruses that are associated with significant expansion of the virome during lentiviral infection. In a first aspect, the invention features isolated polynucleotides including all or a portion of a nucleotide sequence that is at least 70% identical (e.g., at least 71%, 72%, 73%, or 74% identical), at least 75% identical (e.g., at least 76%, 77%, 78%, or 79% identical), at least 80% identical (e.g., at least 81%, 82%, 83%, or 84% identical), at least 85% identical (e.g., at least 86%, 87%, 88%, or 89% identical), at least 90% identical (e.g., at least 91%, 92%, 93%, or 94% identical), at least 95% identical (e.g., at least 96%, 97%, 98%, or 99% identical), or 100% identical to any one of SEQ ID NOs: 1-107, or a reverse complement thereof. In some embodiments, the isolated polynucleotides include a label (e.g., a fluorophore, a hapten, an enzyme, or a radioisotope). The isolated polynucleotides of the invention may include at least 5, 6, 7, 8, 9, 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000, 2000, 3000, 4000, 5000, 6000, 7000 or more contiguous or non-contiguous nucleotides of a reference polynucleotide molecule. In some embodiments, the polynucleotides of the invention are between 10-100 nucleotides in length, more particularly

between 10-30 nucleotides in length (e.g., 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, or 30 nucleotides in length), and can be at least 70% identical (e.g., at least 71%, 72%, 73%, or 74% identical), at least 75% identical (e.g., at least 76%, 77%, 78%, or 79% identical), at least 80% identical (e.g., at least 81%, 82%, 83%, or 84% identical), at least 85% identical (e.g., at least 86%, 87%, 88%, or 89% identical), at least 90% identical (e.g., at least 91%, 92%, 93%, or 94% identical), at least 95% identical (e.g., at least 96%, 97%, 98%, or 99% identical), or 100% identical to any one of SEQ ID NOs: 332-371.

**[0005]** In another aspect, the invention features isolated polypeptides including all or a portion of an amino acid sequence that is at least 70% identical (e.g., at least 71%, 72%, 73%, or 74% identical), at least 75% identical (e.g., at least 76%, 77%, 78%, or 79% identical), at least 80% identical (e.g., at least 81%, 82%, 83%, or 84% identical), at least 85% identical (e.g., at least 86%, 87%, 88%, or 89% identical), at least 90% identical (e.g., at least 91%, 92%, 93%, or 94% identical), at least 95% identical (e.g., at least 96%, 97%, 98%, or 99% identical), or 100% identical to any one of SEQ ID NOs: 108-331. In some embodiments, the isolated polypeptides include a label (e.g., a fluorophore, a hapten, an enzyme, or a radioisotope). The isolated polypeptides of the invention may include at least 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 25, 50, 75, 90, 100, 125, 150, 175, 200, 225, 250, 275, 300, or 350 or more contiguous or non-contiguous amino acids of a reference polypeptide molecule.

**[0006]** In another aspect, the invention features isolated antibodies, or fragments thereof, that are specific for one or more of the isolated polynucleotides or polypeptides of the invention. In some embodiments, the isolated antibodies, or fragments thereof, may be chimeric, human, humanized, or synthetic. In other embodiments, the isolated antibodies, or fragments thereof, may further include a label.

**[0007]** In another aspect, the invention features recombinant expression systems for the production of a protein, or fragment thereof, that is encoded by the polynucleotides of the invention. In some embodiments, the recombinant expression system is an in vitro or an in vivo expression system. In other embodiments, the recombinant expression system further includes a cell (e.g., a bacterial, plant, or mammalian cell). In yet other embodiments, the mammalian cell is a Chinese hamster ovary (CHO) cell.

**[0008]** In another aspect, the invention features recombinant viruses including one or more of the isolated polynucleotides and/or one or more of the isolated polypeptides of the invention. In one preferred embodiment, the viruses further include a genome including a heterologous nucleic acid encoding an antigenic gene product of interest or fragment thereof, or the viruses further include a capsid including a heterologous antigenic gene product of interest or fragment thereof. In another preferred embodiment, the antigenic gene product, or fragment thereof, includes a bacterial, viral, parasitic, or fungal gene product, or fragment thereof. In some embodiments, all or a portion of the recombinant virus is from the viral family Adenoviridae, Parvoviridae, Calciviridae, Papillomaviridae, Picobirnaviridae, Picornaviridae, or Polyomaviridae. In preferred embodiments, the viral family is Adenoviridae or Parvoviridae.

**[0009]** In another aspect, the invention features methods of detecting acquired immune deficiency syndrome (AIDS) and/or AIDS progression in a subject including detecting one or more target nucleotide sequences from a sample of the

subject that specifically hybridize under stringent conditions to one or more of the polynucleotides of the invention, where the detection of an increase in the level of the one or more target nucleotide sequences in the subject, relative to the level of one or more target nucleotide sequences from a control subject, indicates AIDS and/or AIDS progression in the subject.

**[0010]** In another aspect, the invention features methods of diagnosing, or providing a prognostic indicator of, immunodeficiency and/or enteropathy in a subject including detecting one or more target nucleotide sequences from a sample of the subject that specifically hybridize under stringent conditions to one or more of the polynucleotides of the invention, where the detection of an increase in the level of the one or more target nucleotide sequences in the subject, relative to the level of one or more target nucleotide sequences from a control subject, indicates the presence of, or the propensity to develop, immunodeficiency and/or enteropathy in the subject. In some embodiments, the immunodeficiency and/or enteropathy is associated with a lentivirus (e.g., human immunodeficiency virus (HIV) or simian immunodeficiency virus (SIV)).

**[0011]** In yet another aspect, the invention features methods of identifying a therapeutic agent for use in treating immunodeficiency and/or enteropathy in a subject including detecting one or more target nucleotide sequences that specifically hybridize under stringent conditions to one or more of the polynucleotides of the invention from a sample of a subject administered a therapeutically effective amount of a candidate agent, where the detection of a decrease in the level of the one or more target nucleotide sequences in the subject, relative to the level of one or more target nucleotide sequences from the subject prior to administration or a control subject, identifies the candidate agent as the therapeutic agent. In some embodiments, the candidate agent is administered to the subject in a therapeutically effective amount. In other embodiments, the immunodeficiency and/or enteropathy is associated with a lentivirus (e.g., human immunodeficiency virus (HIV) or simian immunodeficiency virus (SIV)).

**[0012]** In other aspects, the invention features nucleic acid-based vaccines including a vector including the polynucleotides of the invention as well as isolated recombinant cells including the polynucleotides of the invention (e.g., all or a portion of a polynucleotide having at least 70%, 80%, 85%, 90%, 95%, 99%, or 100% sequence identity to any one or more of SEQ ID NOs: 1-107 and/or 332-371, or a reverse complement thereof).

**[0013]** In other aspects, the invention features polypeptide-based vaccines including the polypeptides of the invention (e.g., all or a portion of a polypeptide having at least 70%, 80%, 85%, 90%, 95%, 99%, or 100% sequence identity to any one or more of SEQ ID NOs: 108-331).

**[0014]** In any of the methods described herein, the detecting of one or more target nucleotide sequences may include synthesizing cDNA from RNA of the sample.

**[0015]** In any of the embodiments described herein, the one or more target nucleotide sequences are detected by a PCR assay (e.g., a real time PCR (RT-PCR) assay and/or a nested PCR assay).

**[0016]** In any of the embodiments described herein, the sample is a tissue, organ, liquid, or feces sample. In preferred embodiments, the sample is from a mammal, preferably a primate, such as a human.

## DEFINITIONS

**[0017]** The term “antibody” is used in the broadest sense and includes monoclonal antibodies (e.g., full-length or intact monoclonal antibodies), polyclonal antibodies, multivalent antibodies, multispecific antibodies (e.g., bispecific antibodies) so long as they exhibit the desired biological activity and may also include certain antibody fragments (as described in greater detail herein). An antibody typically comprises both “light chains” and “heavy chains.” The light chains of antibodies (immunoglobulins) from any vertebrate species can be assigned to one of two clearly distinct types, called kappa ( $\kappa$ ) and lambda ( $\lambda$ ), based on the amino acid sequences of their constant domains. Depending on the amino acid sequence of the constant domain of their heavy chains, immunoglobulins can be assigned to different classes. There are five major classes of immunoglobulins: IgA, IgD, IgE, IgG, and IgM, and several of these can be further divided into subclasses (isotypes), e.g., IgG1, IgG2, IgG3, IgG4, IgA1, and IgA2. The heavy chain constant domains that correspond to the different classes of immunoglobulins are called  $\alpha$ ,  $\delta$ ,  $\epsilon$ ,  $\gamma$ , and  $\mu$ , respectively. The subunit structures and three-dimensional configurations of different classes of immunoglobulins are well known.

**[0018]** “Antibody fragments” of “fragments” comprise only a portion of an intact antibody, wherein the portion preferably retains at least one, preferably most or all, of the functions normally associated with that portion when present in an intact antibody. Examples of antibody fragments include Fab, Fab', F(ab')<sub>2</sub>, and Fv fragments (e.g., single-chain variable fragments (scFv)); diabodies; linear antibodies; single-chain antibody molecules; and multispecific antibodies formed from antibody fragments. Papain digestion of antibodies produces two identical antigen-binding fragments, called “Fab” fragments, each with a single antigen-binding site, and a residual “Fc” fragment, whose name reflects its ability to crystallize readily. Pepsin treatment yields an F(ab')<sub>2</sub> fragment that has two antigen-combining sites and is still capable of cross-linking antigen. In one embodiment, an antibody fragment comprises an antigen binding site of the intact antibody and thus retains the ability to bind antigen. In another embodiment, an antibody fragment, for example one that comprises the Fc region, retains at least one of the biological functions normally associated with the Fc region when present in an intact antibody, such as FcRn binding, antibody half life modulation, ADCC function, ADCVI function, and complement binding. In one embodiment, an antibody fragment is a monovalent antibody that has an in vivo half life substantially similar to an intact antibody. For example, such an antibody fragment may comprise an antigen binding arm linked to an Fc sequence capable of conferring in vivo stability to the fragment.

**[0019]** By “capsid” is meant a protein shell or coat of a virus which often adopts a helical or icosahedral structure. The capsid of an adenovirus, for example, adopts an icosahedral structure and consists of three major structural proteins: hexon, penton, and fiber proteins. The capsid encloses the genetic material of the virus.

**[0020]** Throughout this specification and claims, the word “comprise,” or variations such as “comprises” or “comprising,” will be understood to imply the inclusion of a stated integer or group of integers but not the exclusion of any other integer or group of integers.

**[0021]** By “enteropathy” is meant damage to the intestinal epithelium, commonly associated with lentiviral infection

(e.g., human immunodeficiency virus (HIV) infection in humans), which can result in intestinal leakage associated with increased serum LPS binding protein (LBP) levels and systemic immune activation. Enteritis, inflammation of the intestinal epithelium, is a type of enteropathy.

**[0022]** By “gene product” is meant to include mRNAs or other nucleic acids (e.g., microRNAs) transcribed from a gene as well as polypeptides translated from those mRNAs.

**[0023]** By “heterologous nucleic acid molecule” is meant any exogenous nucleic acid molecule that can be incorporated into, for example, a virus or expression system of the invention for subsequent expression of a gene product of interest or fragment thereof encoded by the heterologous nucleic acid molecule. In a preferred embodiment, the heterologous nucleic acid molecule encodes an antigenic gene that is of bacterial, viral, parasitic, or fungal origin (e.g., a nucleic acid molecule encoding the HIV Gag, Pol, Env, Nef, Tat, Rev, Vif, Vpr, or Vpu gene product, or fragment thereof). The heterologous nucleic acid molecule is one that is not normally associated with the other nucleic acid molecules found in virus or expression system.

**[0024]** By “immunodeficiency” is meant a compromised immune system of a subject relative to that of a control, whereby the compromise of the immune system can be measured by a decrease in the levels of CD4 T cells, B cells, plasma cells, antibodies, or neutrophil granulocytes of the subject relative to that of the control.

**[0025]** By “isolated” is meant separated, recovered, or purified from a component of its natural environment.

**[0026]** A “label” refers to a molecular moiety or compound that is detected or leads to a detectable signal. A label may be joined directly or indirectly to a polynucleotide, polypeptide, or a probe thereof. Direct labeling can occur through bonds or interactions that link the label to the probe, including covalent bonds or non-covalent interactions, e.g. hydrogen bonds, hydrophobic and ionic interactions, or formation of chelates or coordination complexes. Indirect labeling can occur through use of a bridging moiety or linker (e.g., antibody or additional oligomer), which is either directly or indirectly labeled, and which may amplify the detectable signal. Labels include any detectable moiety, such as a fluorophore, hapten, enzyme, radioisotope, enzyme substrate, reactive group, chromophore (e.g., a dye, a particle, or a bead that imparts detectable color), or luminescent compound (e.g., bioluminescent, phosphorescent, or chemiluminescent labels). A “radioisotope” can be any radioisotope known to skilled artisans, such as,  $^3\text{H}$ ,  $^{14}\text{C}$ ,  $^{32}\text{P}$ ,  $^{33}\text{P}$ ,  $^{35}\text{S}$ , or  $^{125}\text{I}$ . A “fluorophore” can be any fluorophore known to skilled artisan, for example, a fluorescein, a rhodamine, a coumarin, an indocyanine, or a green fluorescent protein (GFP) or variant thereof (e.g., a red fluorescent protein (RFP)). An enzyme can be any enzyme for which a suitable substrate is available, such as, for example, alkaline phosphatase, a horseradish peroxidase, or a chloramphenicol acetyltransferase. A suitable substrate is a substrate that, when contacted by an enzyme, produces a product that is detectable by methods known to skilled artisans. For example, the substrate can be a chromogenic substrate (e.g., p-dinitrophenyl phosphate as a substrate for alkaline phosphatase or diamminobenzidine as a substrate for horseradish peroxidase), a fluorogenic substrate (e.g., 5-amino-2,3-dihydrophthalazine-1,4-dione (luminol) for horseradish peroxidase or disodium 3-(4-methoxy Spiro{1,2-dioxetane-3,2'-(5'chloro)tricyclo[3.3.1.1.3,7]decan}-4-yl)phenyl phosphate

for alkaline phosphatase). A “hapten” can be any hapten for which a probe is available, such as biotin, streptavidin, or digoxigenin.

**[0027]** By “portion” is meant a part of a whole. A portion may comprise at least 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, or 95% of the entire length of a polynucleotide or polypeptide sequence region. For polynucleotides, for example, a portion may include at least 5, 6, 7, 8, 9, 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000, 2000, 3000, 4000, 5000, 6000, 7000 or more contiguous nucleotides of a reference polynucleotide molecule. For polypeptides, for example, a portion may include at least 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 25, 50, 75, 90, 100, 125, 150, 175, 200, 225, 250, 275, 300, or 350 or more contiguous amino acids of a reference polypeptide molecule.

**[0028]** By “recombinant,” with respect to an expression system or virus, is meant an expression system or virus that has been manipulated in vitro. For example, an expression system or virus which includes a heterologous nucleic acid sequence, such as a sequence encoding an antigenic gene product, introduced using recombinant nucleic acid techniques.

**[0029]** By “sample” is meant any biological substance obtained from a subject, such as a biological feces (stool), fluid, tissue, or organ sample. A biological fluid sample can be, without limitation, a blood sample, a plasma sample, a serum sample, a cerebrospinal fluid sample, a urine sample, or a saliva sample.

**[0030]** By “sequence identity” or “sequence similarity” is meant that the identity or similarity between two or more amino acid sequences, or two or more nucleotide sequences, is expressed in terms of the identity or similarity between the sequences. Sequence identity can be measured in terms of “percentage (%) identity,” wherein the higher the percentage, the more identity shared between the sequences. Sequence similarity can be measured in terms of percentage similarity (which takes into account conservative amino acid substitutions); the higher the percentage, the more similarity shared between the sequences. Homologs or orthologs of nucleic acid or amino acid sequences possess a relatively high degree of sequence identity similarity when aligned using standard methods. Sequence identity may be measured using sequence analysis software on the default setting (e.g., Sequence Analysis Software Package of the Genetics Computer Group, University of Wisconsin Biotechnology Center, 1710 University Avenue, Madison, Wis. 53705). Such software may match similar sequences by assigning degrees of homology to various substitutions, deletions, and other modifications.

**[0031]** By “specifically hybridizes” is meant hybridization, under stringent hybridization conditions, of a first polynucleotide (e.g., a probe or primer) to a second polynucleotide (e.g., a target sequence) to a detectably greater degree than hybridization of the first polynucleotide to non-target polynucleotide sequences and/or to the substantial exclusion of non-target polynucleotide sequences. Selectively hybridizing sequences have at least 70% sequence identity, at least 80% sequence identity, at least 90% sequence identity, or 100% sequence identity (e.g., complementary) with each other.

**[0032]** The term “stringent conditions” refers to conditions under which a probe will hybridize to its target sequence to a detectably greater degree than to other sequences. Stringent conditions are sequence-dependent and will be different in different circumstances. Longer sequences hybridize specifi-

cally at higher temperatures. Generally, stringent conditions are selected to be about 5° C. lower than the thermal melting point ( $T_m$ ) for the specific sequence at a defined ionic strength and pH. The  $T_m$  is the temperature (under defined ionic strength and pH) at which 50% of a complementary target sequence hybridizes to a perfectly matched probe. Typically, stringent conditions will be those in which the salt concentration is less than about 1.0 M Na ion, typically about 0.01 to 1.0 M Na ion concentration (or other salts) at pH 7.0 to 8.3 and the temperature is at least about 30° C. for short probes (e.g., 10 to 50 nucleotides) and at least about 60° C. for long probes (e.g., greater than 50 nucleotides). Stringent conditions may also be achieved with the addition of destabilizing agents, such as formamide.

**[0033]** A “subject” is a vertebrate, such as a mammal (e.g., primates and humans). Mammals also include, but are not limited to, farm animals (such as cows), sport animals, pets (such as cats, dogs, and horses), mice, and rats.

**[0034]** By “therapeutically effective amount” is meant an amount of a therapeutic agent that alone, or together with one or more additional (optional) therapeutic agents, produces beneficial or desired results upon administration to a mammal. The therapeutically effective amount depends upon the context in which the therapeutic agent is applied. For example, in the context of administering a composition including a therapeutic agent, the therapeutically effective amount of the composition is an amount sufficient to achieve a reduction in the level of an infectious virus, such as HIV or SIV (e.g., as measured by a stabilization or an increase in CD4 T cell count relative to a control), and/or a reduction in the level of enteropathy (e.g., as measured by a decrease in serum LBP levels relative to a control) as compared to a response obtained without administration of the composition, and/or to prevent the propagation of an infectious virus (e.g., HIV) in a subject (e.g., a human) having an increased risk of viral infection. Ideally, a therapeutically effective amount provides a therapeutic effect without causing a substantial cytotoxic effect in the subject. In general, a therapeutically effective amount of a composition administered to a subject (e.g., a human subject) will vary depending upon a number of factors associated with that subject, for example the overall health of the subject, the condition to be treated, or the severity of the condition. A therapeutically effective amount of a composition can be determined by varying the dosage of the product and measuring the resulting therapeutic response.

**[0035]** As used herein, and as well understood in the art, “treatment” is an approach for obtaining beneficial or desired results, such as clinical results. Beneficial or desired results can include, but are not limited to, alleviation or amelioration of one or more symptoms or conditions; diminishment of extent of disease, disorder, or condition; stabilization (i.e., not worsening) of a state of disease, disorder, or condition; prevention of spread of disease, disorder, or condition; delay or slowing the progress of the disease, disorder, or condition; amelioration or palliation of the disease, disorder, or condition; and remission (whether partial or total), whether detectable or undetectable. “Palliating” a disease, disorder, or condition means that the extent and/or undesirable clinical manifestations of the disease, disorder, or condition are lessened and/or time course of the progression is slowed or lengthened, as compared to the extent or time course in the absence of treatment.

**[0036]** Other features and advantages of the invention will be apparent from the following Detailed Description, the drawings, and the claims.

#### BRIEF DESCRIPTION OF THE DRAWINGS

**[0037]** FIG. 1A is a graph showing the taxonomic distribution of sequences identified in feces of pathogenic SIV-infected (SIV+) and uninfected (SIV-) control rhesus monkeys housed at the NEPRC 24 weeks after intrarectal infection with SIVmac251. The flanking doughnut chart displays the averaged values per kingdom for SIV+ or SIV- monkeys.

**[0038]** FIG. 1B is a graph showing the taxonomic distribution of sequences identified in feces of pathogenic SIV-infected (SIV+) and uninfected (SIV-) control rhesus monkeys described in FIG. 1A housed at the NEPRC 64 weeks after SIV infection. \* indicates euthanized for progressive AIDS 24 to 64 weeks after SIV infection. The flanking doughnut chart displays the averaged values per kingdom for SIV+ or SIV- monkeys.

**[0039]** FIG. 1C is a graph showing the taxonomic distribution of sequences identified in feces of pathogenic SIV-infected (SIV+) and uninfected (SIV-) control rhesus monkeys housed at the TNPRC 23-64 weeks after intravaginal infection with SIVmac251. The flanking doughnut chart displays the averaged values per kingdom for SIV+ or SIV- monkeys.

**[0040]** FIG. 1D is a graph showing taxonomic distribution of sequences identified in feces of non-pathogenic SIV-infected (SIV+) and control (SIV-) vervet African green monkeys housed at the NIH at least three years after intravenous infection with SIVagm90, SIVagmVer1, or after natural infection in the wild. The flanking doughnut chart displays the averaged values per kingdom for SIV+ or SIV- monkeys.

**[0041]** FIG. 1E is a graph showing the taxonomic distribution of sequences identified in feces of non-pathogenic SIV-infected (SIV+) and control (SIV-) sabaues African green monkeys housed at the NEPRC and infected intravenously with SIVagmMJ8, SIVagm9315BR, or uninfected controls. The flanking doughnut chart displays the averaged values per kingdom for SIV+ or SIV- monkeys.

**[0042]** FIG. 2A is a graph showing the quantitation of sequences from different kingdoms of life identified in the feces of pathogenic SIV-infected and control rhesus monkeys housed at the NEPRC 24 weeks after SIV infection. The nature of SIV infection is as defined in the FIG. 1A.

**[0043]** FIG. 2B is a graph showing the quantitation of sequences from different kingdoms of life identified in the feces of pathogenic SIV-infected and control rhesus monkeys housed at the NEPRC 64 weeks after SIV infection. The nature of SIV infection is as defined in the FIG. 1B.

**[0044]** FIG. 2C is a graph showing the quantitation of sequences from different kingdoms of life identified in the feces of pathogenic SIV-infected and control rhesus monkeys housed at the TNPRC. The nature of SIV infection is as defined in the FIG. 1C.

**[0045]** FIG. 2D is a graph showing the quantitation of sequences from different kingdoms of life identified in the feces of nonpathogenic SIV-infected and control vervet African green monkeys housed at the NIH. The nature of SIV infection is as defined in the FIG. 1D.

**[0046]** FIG. 2E is a graph showing the quantitation of sequences from different kingdoms of life identified in the feces of nonpathogenic SIV-infected and control sabaues African green monkeys housed at the NEPRC. The nature of SIV infection is as defined in the FIG. 1E.

**[0047]** FIG. 3A is a graph showing SIV RNA levels in animals in the NEPRC cohort.

**[0048]** FIG. 3B is a graph showing CD4 T cell (CD4) numbers in animals in the NEPRC cohort.

**[0049]** FIG. 3C is a graph showing serum LPS binding protein (LBP) levels in animals in the NEPRC cohort.

**[0050]** FIG. 4A is a chart showing the distribution of virus sequences present in the feces of pathogenic SIV-infected and control rhesus monkeys housed at the NEPRC 24 weeks after SIV infection. "Mammalian" indicates that sequences were most closely related to viruses that infect mammals. Viruses infecting non-mammals are referred to as "other." "Unclassified viruses" includes all unclassified viruses, e.g., Chronic bee paralysis virus, Chimpanzee stool associated circular ssDNA virus, Circovirus-like genome RW-C, Circovirus-like genome CB-A, and Rodent stool-associated circular genome virus.

**[0051]** FIG. 4B is a chart showing the distribution of virus sequences present in the feces of pathogenic SIV-infected and control rhesus monkeys housed at the NEPRC 64 weeks after SIV infection. \* indicates euthanized for progressive AIDS between 24 and 64 weeks after SIV infection. Virus classifications as described in FIG. 4A.

**[0052]** FIG. 4C is a chart showing the distribution of virus sequences present in the feces of pathogenic SIV-infected and control rhesus monkeys housed at the TNPRC. Virus classifications as described in FIG. 4A.

**[0053]** FIG. 4D is a chart showing the distribution of virus sequences present in the feces of non-pathogenic SIV-infected and control vervet African green monkeys housed at the NIH. Virus classifications as described in FIG. 4A.

**[0054]** FIG. 4E is a chart showing the distribution of virus sequences present in the feces of non-pathogenic SIV-infected and control sabaes African green monkeys housed at the NEPRC. (C) Viruses present in feces of pathogenic SIV-infected and control rhesus monkeys housed at the TNPRC. Virus classifications as described in FIG. 4A.

**[0055]** FIG. 4F is a graph showing the average number of picornavirus sequences, after normalization for analysis using MEGAN, detected in the indicated cohorts of SIV-infected (+) and control (-) rhesus monkeys.

**[0056]** FIG. 5A are schematic diagrams showing the assembled viral contigs (in gray) from newly identified WUHARV Caliciviruses 1 (animal 39), 2 (from an animal not included in the cohort), and 3 (animal 39) compared to Tulane calicivirus (black bar). Calicivirus 1 contig 1 derived from 879 sequences, length=6578 bp; Calicivirus 2 contig 1 derived from 16 sequences, length=812 bp; Calicivirus 2 contig 2 assembled from 120 sequences, length=5083 bp; Calicivirus 3 contig 1 assembled from 14 sequences, length=750 bp; Calicivirus 3 contig 2 assembled from 67 sequences, length=2111 bp; Calicivirus 3 contig 3 assembled from 41 sequences, length=832 bp; Calicivirus 3 contig 4 assembled from 38 sequences, length=1273 bp. Animal numbers refer to the monkeys in FIG. 1A. \*Indicates the percentage nucleotide identity over the designated length of the best aligned homologous region (indicated by double headed arrow) compared to the most closely related genome indicated in the black bar.

**[0057]** FIG. 5B are schematic diagrams showing the assembled viral contigs (in gray) from newly identified WUHARV Parvovirus 1 (animal 39) and 2 (animal 35) compared with the sequence of canine or mouse parvovirus 4a (black bars), as indicated. Parvovirus 1 contig 1 assembled

from 375 sequences, length=4905 bp; Parvovirus 2 contig 1 representing 1 sequence, length=470 bp; Parvovirus 2 contig 2 assembled from 6 sequences, length=690 bp. Animal numbers refer to the monkeys in FIG. 1A. \*Indicates the percentage nucleotide identity over the designated length of the best aligned homologous region (indicated by double headed arrow) compared to the most closely related genome indicated in the black bar.

**[0058]** FIG. 5C are schematic diagrams showing the assembled viral contigs (in gray) from newly identified WUHARV Enterovirus 1 (animal 41), 2 (animal 39) and 3 (animal 33) compared with the sequence of Simian enterovirus SV19 (black bar). Enterovirus 1 assembled from 1084 sequences, length=7273 bp; Enterovirus 2 assembled from 758 sequences, length=7128 bp; Enterovirus 3 assembled from 406 sequences, length=6962 bp. Animal numbers refer to the monkeys in FIG. 1A. \*Indicates the percentage nucleotide identity over the designated length of the best aligned homologous region (indicated by double headed arrow) compared to the most closely related genome indicated in the black bar.

**[0059]** FIG. 5D are schematic diagrams showing the assembled viral contigs (in gray) from newly identified WUHARV Sapelovirus 1 (animal 42), 2 (animal 41) and 3 (animal 37) compared with the sequence of Simian Sapelovirus 1 strain 2383 (black bar). Sapelovirus 1 assembled from 3081 sequences, length=8059 bp; Sapelovirus 2 assembled from 2711 sequences, length=8025 bp; Sapelovirus 3 assembled from 380 sequences, length=6872 bp. Animal numbers refer to the monkeys in FIG. 1A. \*Indicates the percentage nucleotide identity over the designated length of the best aligned homologous region (indicated by double headed arrow) compared to the most closely related genome indicated in the black bar.

**[0060]** FIG. 5E is a chart showing the presence of viral sequences as detected by PCR using virus-specific primers (Table 1). Numbers below the chart refer to the animals in FIG. 1A. "a" refers to lack of detection of a virus likely due to the presence of a divergent virus; "b" refers to lack of detection of a virus for unknown reasons; and "c" refers to detection of virus sequences in serum samples taken at the time of euthanasia for AIDS.

**[0061]** FIG. 6A are schematic diagrams showing the assembled viral contigs (in gray) from newly identified WUHARV Adenovirus 1 (animal #40) compared to the known virus Simian adenovirus 1 strain ATCCVR-195 (black bar). These contigs were assembled from 1308 sequences. Animal numbers refer to the monkeys in FIG. 1A. \*Indicates the percentage nucleotide identity over the designated length of the best aligned homologous region (indicated by double headed arrow) compared to the most closely related genome indicated in the black bar.

**[0062]** FIG. 6B is an agarose gel showing PCR confirmation of WUHARV Adenovirus 1 during amplification, plaque purification, and cesium chloride gradient purification. The three PCR products for each sample (lanes 2-19) were derived from primers 4302c3f and 4302c3r, 4302c18f and 4302c18r, and 4302c1f and 4302c1 r, respectively (Table 1). Lane 1 is a molecular weight ladder.

**[0063]** FIG. 6C are images showing representative histopathology (top panels) and adenovirus immunohistochemistry (IHC) (bottom panels) for animal #23. Adenovirus infection was associated with villous atrophy and fusion (i) and sloughed epithelial cells that contained intranuclear adenovi-

ral inclusions (arrows in (ii)). Adenovirus antigen could be localized to villous tip epithelium by immunohistochemistry (brown color of DAB chromagen, Mayer's counterstain; (iii) and (iv)). Scale bars in (i) and (iii) are 0.5 mm. Scale bars in (ii) and (iv) are 200  $\mu$ m.

**[0064]** FIG. 6D are images showing representative histopathology (top panels) and adenovirus immunohistochemistry (IHC) (bottom panels) for animal #27. Adenovirus infection was associated with villous atrophy and fusion (i) and sloughed epithelial cells that contained intranuclear adenoviral inclusions (arrows in (ii)). Adenovirus antigen could be localized to villous tip epithelium by immunohistochemistry (brown color of DAB chromagen, Mayer's counterstain; (iii) and (iv)). Scale bars in (i) and (iii) are 0.5 mm. Scale bars in (ii) and (iv) are 200  $\mu$ m.

**[0065]** FIG. 7A is a diagram showing the neighbor-joining phylogenetic analysis of the predicted full-length non-structural polyprotein of WUHARV Calicivirus 1.

**[0066]** FIG. 7B is a diagram showing the neighbor-joining phylogenetic analysis of the predicted non-structural protein of WUHARV Parvovirus 1.

**[0067]** FIG. 7C is a diagram showing the neighbor-joining phylogenetic analysis of the full genome of WUHARV Enteroviruses 1, 2, and 3, and WUHARV Sapeloviruses 1 and 2.

**[0068]** FIG. 8A is a rank abundance plot for SIV- animals constructed using both 16S rDNA sequencing from a previous study performed from TNPRC (McKenna et al., *PLoS Pathog.* 4: e20 (2008)) and our next-generation sequencing (NGS) data from TNPRC (Table 2; FIG. 9C). Samples for 16S rDNA sequencing were obtained over a period of months in 1996, while samples for this study were obtained in 2011. Circles indicate 16S rDNA sequencing data from McKenna et al.; squares indicate sequencing data from this study.

**[0069]** FIG. 8B is a rank abundance plot for SIV+ animals constructed using both 16S rDNA sequencing from a previous study performed from TNPRC (McKenna et al., *PLoS Pathog.* 4: e20 (2008)) and our next-generation sequencing (NGS) data from TNPRC (Table 2; FIG. 9C). Samples for 16S rDNA sequencing were obtained over a period of months in 1996, while samples for this study were obtained in 2011. Circles indicate 16S rDNA sequencing data from McKenna et al.; squares indicate sequencing data from this study.

**[0070]** FIG. 8C are graphs showing species accumulation (left panel), Shannon's diversity (middle panel), and Pielou's evenness (right panel) for SIV-infected and control monkeys housed at NEPRC for 24 weeks. The species accumulation curve was constructed for SIV-infected (red) and uninfected control (blue) rhesus monkeys by quantifying the average number of bacterial families identified as additional animals were added to the analysis. The corresponding Shannon's diversity and Pielou's evenness ranges were calculated for equivalent sample numbers based on the minimum sample number between SIV+ and SIV- animals. When this minimum number was less than the maximum number of animals, 100 random samples with replacement were used to determine the sample mean. Differences between means were assessed using an unpaired Student's t-test. The nature of SIV infection is as defined in FIGS. 1A-1D.

**[0071]** FIG. 8D are graphs showing species accumulation (left panel), Shannon's diversity (middle panel), and Pielou's evenness (right panel) for SIV-infected and control monkeys housed at NEPRC for 64 weeks. The species accumulation curve and corresponding Shannon's diversity and Pielou's

evenness ranges were calculated as described for FIG. 8C. The nature of SIV infection is as defined in FIGS. 1A-1D.

**[0072]** FIG. 8E are graphs showing species accumulation (left panel), Shannon's diversity (middle panel), and Pielou's evenness (right panel) for SIV-infected and control monkeys housed at TNPRC 23-64 weeks after intravaginal infection with SIVmac251. The species accumulation curve and corresponding Shannon's diversity and Pielou's evenness ranges were calculated as described for FIG. 8C. The nature of SIV infection is as defined in FIGS. 1A-1D.

**[0073]** FIG. 8F are graphs showing species accumulation (left panel), Shannon's diversity (middle panel), and Pielou's evenness (right panel) for SIV-infected and control vervet African green monkeys housed at the NIH after intravenous infection with SIVagm90, SIVagmVer1, or after natural infection in the wild. The species accumulation curve and corresponding Shannon's diversity and Pielou's evenness ranges were calculated as described for FIG. 8C. The nature of SIV infection is as defined in FIGS. 1A-1D.

**[0074]** FIG. 8G is a graph showing species accumulation for SIV-infected and control sabaecus African green monkeys housed at NEPRC and infected intravenously with SIVagm-MJB or SIVagm9315BR. The species accumulation curve was calculated as described for FIG. 8C. Based on the lack of comparable family richness between SIV-infected and control animals in this cohort, we do not report diversity or evenness.

**[0075]** FIG. 9A is a heatmap displaying the number of sequences assigned to specific bacterial families for each individual pathogenic SIV-infected and control rhesus monkey housed at the NEPRC 24 weeks after SIV infection. The nature of SIV infection is as defined for FIG. 1A.

**[0076]** FIG. 9B is a heatmap displaying the number of sequences assigned to specific bacterial families for each individual pathogenic SIV-infected and control rhesus monkey housed at the NEPRC 64 weeks after SIV infection. The nature of SIV infection is as defined for FIG. 1B.

**[0077]** FIG. 9C is a heatmap displaying the number of sequences assigned to specific bacterial families for each individual pathogenic SIV-infected and control rhesus monkeys housed at the TNPRC. The nature of SIV infection is as defined for FIG. 1C.

**[0078]** FIG. 9D is a heatmap displaying the number of sequences assigned to specific bacterial families for each individual nonpathogenic SIV-infected and control vervet African green monkeys housed at the NIH. The nature of SIV infection is as defined for FIG. 1D.

#### DETAILED DESCRIPTION

**[0079]** The present invention relates to the discovery that pathogenic SIV infection is associated with a significant and unexpected expansion of the enteric virome, as detected using next generation sequencing (NGS) of RNA plus DNA. We documented a remarkable number of differences in the fecal virome between pathogenically SIV-infected monkeys, uninfected control monkeys, and monkeys infected with non-pathogenic SIV. These findings included increases in viral sequences, the presence of novel viruses, the association of unsuspected adenovirus infection with intestinal disease and enteric epithelial pathology, and viremia with enteric parvoviruses in advanced AIDS. At least 32 new viruses were detected from genera that cause diseases in mammalian hosts including adenoviruses, caliciviruses, parvoviruses, picornaviruses, and polyomaviruses (see, for example, Table 3 or

FIG. 5 for a summary of the identified viruses). Our assignment of viral sequences to new viruses was conservative, and thus additional sequencing may detect additional viruses in the enteric virome in SIV-infected animals.

**[0080]** Application of standard diagnostic approaches, such as PCR or culture, would not have identified the breadth of divergent viruses detected here, and therefore would have underestimated both the potential causes of enteritis or systemic viral infection and the diversity of antigens which might contribute to enteropathy and immune activation. Our findings show that the nature of the enteric virome can be used as a prognostic indicator of HIV progression. The nature of the enteric virome may also contribute to AIDS pathogenesis by damaging the intestinal epithelium to allow access of microbes, PAMPs, and viral antigens into tissues and the circulation to activate the immune system and stimulate lentivirus replication.

**[0081]** These data challenge the notion that abnormalities in the intestinal tract in pathogenic SIV-infected primates are due to direct effects of SIV or indirect effects of SIV on immune responses to enteric bacteria (Sandler et al., *J. Infect. Dis.* 203: 780-790 (2011)). Instead, immunocompromise during lentivirus infection appears to be associated with significant expansion of the enteric virome, which results in damage to the intestine, as shown for adenoviruses in the present study.

**[0082]** Such damage could provide access for bacterial PAMPs, or as shown here enteric viruses, into tissues and the circulation. It is already recognized that bacterial and viral contributions to intestinal pathology are not independent of each other. Clear synergies between the virome, bacteria, and host genes have been documented in murine systems (Bloom et al., *Cell Host Microbe* 9: 390-403 (2011); Cadwell et al., *Cell* 141: 1135-1145 (2010); Virgin et al., *Cell* 147: 44-56 (2011)). Importantly, it is not clear how bacterial PAMPs would explain the T cell activation characteristic of the systemic immune activation associated with AIDS progression. Our data suggest that T and B cell activation might be due to immune responses to unexpected viral antigens, as for example the parvovirus we detected in the circulation of a subset of animals. Unsuspected viral infections might also contribute to the high levels of IFN- $\alpha$  noted in the circulation of untreated AIDS patients. Searching for virus-specific T cell responses requires knowledge of the sequence of the viral proteins present, indicating the importance of sequencing the virome to define potential antigens that might drive immune activation in lentivirus-infected hosts.

#### Polynucleotides of the Invention

**[0083]** As a first aspect, the invention provides polynucleotide sequences related to previously undiscovered viruses of the enteric virome. The isolated polynucleotides may include all or a portion of a nucleotide sequence that is at least 70% identical (e.g., at least 71%, 72%, 73%, or 74% identical), at least 75% identical (e.g., at least 76%, 77%, 78%, or 79% identical), at least 80% identical (e.g., at least 81%, 82%, 83%, or 84% identical), at least 85% identical (e.g., at least 86%, 87%, 88%, or 89% identical), at least 90% identical (e.g., at least 91%, 92%, 93%, or 94% identical), at least 95% identical (e.g., at least 96%, 97%, 98%, or 99% identical), or 100% identical to any one of SEQ ID NOs: 1-107, or a reverse complement thereof. In some embodiments, the isolated polynucleotides include a label (e.g., a fluorophore, a hapten, an enzyme, or a radioisotope). The isolated polynucleotides

of the invention may include at least 5, 6, 7, 8, 9, 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000, 2000, 3000, 4000, 5000, 6000, 7000 or more contiguous or non-contiguous nucleotides of a reference polynucleotide molecule. In some embodiments, the polynucleotides of the invention are between 10-100 nucleotides in length, more particularly between 10-30 nucleotides in length (e.g., 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, or 30 nucleotides in length), and can be at least 70% identical (e.g., at least 71%, 72%, 73%, or 74% identical), at least 75% identical (e.g., at least 76%, 77%, 78%, or 79% identical), at least 80% identical (e.g., at least 81%, 82%, 83%, or 84% identical), at least 85% identical (e.g., at least 86%, 87%, 88%, or 89% identical), at least 90% identical (e.g., at least 91%, 92%, 93%, or 94% identical), at least 95% identical (e.g., at least 96%, 97%, 98%, or 99% identical), or 100% identical to any one of SEQ ID NOs: 332-371. SEQ ID NOs: 332-371 (see, for example, Table 1) disclose primers that can be utilized in a PCR assay to screen for the presence of the viruses.

#### Polypeptides of the Invention

**[0084]** In another aspect, the invention features isolated polypeptides including all or a portion of an amino acid sequence that is at least 70% identical (e.g., at least 71%, 72%, 73%, or 74% identical), at least 75% identical (e.g., at least 76%, 77%, 78%, or 79% identical), at least 80% identical (e.g., at least 81%, 82%, 83%, or 84% identical), at least 85% identical (e.g., at least 86%, 87%, 88%, or 89% identical), at least 90% identical (e.g., at least 91%, 92%, 93%, or 94% identical), at least 95% identical (e.g., at least 96%, 97%, 98%, or 99% identical), or 100% identical to any one of SEQ ID NOs: 108-331. In some embodiments, the isolated polypeptides include a label (e.g., a fluorophore, a hapten, an enzyme, or a radioisotope). The isolated polypeptides of the invention may include at least 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 25, 50, 75, 90, 100, 125, 150, 175, 200, 225, 250, 275, 300, or 350 or more contiguous or non-contiguous amino acids of a reference polypeptide molecule.

#### Antibodies, Recombinant Expression Systems, and Viruses of the Invention

**[0085]** The invention features isolated antibodies, or fragments thereof, that are specific for one or more of the isolated polynucleotides or polypeptides of the invention. The isolated antibodies, or fragments thereof, may be chimeric, human, humanized, or synthetic, and may further include a label.

**[0086]** In another aspect, the invention features recombinant expression systems for the production of a protein, or fragment thereof, that is encoded by the polynucleotides of the invention. The recombinant expression system may be an in vitro or an in vivo expression system and may further include a cell. The cell may be a bacterial cell (e.g., an *E. coli* cell), a plant cell, or a mammalian cell (e.g., a Chinese hamster ovary (CHO) cell).

**[0087]** In yet another aspect, the invention features recombinant viruses including one or more of the isolated polynucleotides and/or one or more of the isolated polypeptides of the invention. In one preferred embodiment, the viruses further include a genome including a heterologous nucleic acid encoding an antigenic gene product of interest or fragment thereof, or the viruses further include a capsid including a heterologous antigenic gene product of interest or fragment



thereof. The antigenic gene product, or fragment thereof, may include a bacterial, viral, parasitic, or fungal gene product, or fragment thereof. Non-limiting examples of bacterial gene products, or fragments thereof, include 10.4, 85A, 85B, 86C, CFP-10, Rv3871, and ESAT-6 gene products, or fragments thereof, of *Mycobacterium*; O, H, and K antigens, or fragments thereof, of *E. coli*; and protective antigen (PA), or fragments thereof, of *Bacillus anthracis*. Non-limiting examples of viral gene products, or fragments thereof, include Gag, Pol, Nef, Tat, Rev, Vif, Vpr, or Vpu, or fragments thereof, of HIV and other retroviruses; 9D antigen, or fragments thereof, of HSV; Env, or fragments thereof, of all envelope protein-containing viruses. Non-limiting examples of parasitic gene products, or fragments thereof, include circumsporozoite (CS) protein, gamete surface proteins Pfs230 and Pfs4845, and Liver Specific Antigens 1 or 3 (LSA-1 or LSA-3), or fragments thereof, of *Plasmodium falciparum*. Non-limiting examples of fungal gene products, or fragments thereof, include any cell wall mannoprotein (e.g., Afmp1 of *Aspergillus fumigatus*) or surface-expressed glycoprotein (e.g., SOWgp of *Coccidioides immitis*). In some embodiments, all or a portion of the recombinant virus is from the viral family Adenoviridae, Parvoviridae, Calciviridae, Papillomaviridae, Picobinaviridae, Picornaviridae, or Polyomaviridae. In preferred embodiments, the viral family is Adenoviridae or Parvoviridae. For example, in some embodiments, all or a portion of the recombinant virus of the invention may be from WUHARV Adenovirus 1 and have all or a portion of a nucleotide sequence that is at least 70% identical to any one of SEQ ID NOs: 1-13 and/or express all or a portion of a polypeptide sequence that is at least 70% identical to any one of SEQ ID NOs: 108-163. In some embodiments, the recombinant virus of the invention may be from WUHARV Adenovirus 2 or 3 and have all or a portion of a nucleotide sequence that is at least 70% identical to any one of SEQ ID NOs: 14-54 and/or express all or a portion of a polypeptide sequence that is at least 70% identical to any one of SEQ ID NOs: 164-256. In some embodiments, the recombinant virus of the invention may be from WUHARV Adenovirus 4 and have all or a portion of a nucleotide sequence that is at least 70% identical to SEQ ID NO: 55 or SEQ ID NO: 56 and/or express all or a portion of a polypeptide sequence that is at least 70% identical to SEQ ID NO: 257 or SEQ ID NO: 258. In some embodiments, the recombinant virus of the invention may be from WUHARV Adenovirus 5 and have all or a portion of a nucleotide sequence that is at least 70% identical to any one of SEQ ID NOs: 57-69 and/or express all or a portion of a polypeptide sequence that is at least 70% identical to any one of SEQ ID NOs: 259-277.

#### Detection of Acquired Immune Deficiency Syndrome (AIDS) or AIDS Progression

**[0088]** Discovery of the expansion of the enteric virome in nonhuman primates infected with pathogenic SIV, but not with non-pathogenic SIV, has profound implications for understanding AIDS pathogenesis in these animals and suggests a similar expansion of the enteric virome in human AIDS. Our data are consistent with a model in which immunosuppression results in increased levels of enteric viral infection which, in a feed-forward manner, contributes to AIDS via damage to the intestinal mucosa and induction of systemic immune activation that accelerates AIDS progression. This study shows the pathogenetic potential of the enteric virome, as exemplified by animals with enteritis asso-

ciated with adenovirus infection or parvovirus viremia. By sequencing both RNA and DNA and by using metagenomic approaches, rather than focusing on bacterial 16S rDNA analysis, we have documented a new set of viruses associated with clinical AIDS progression in rhesus monkeys. Since these viruses include many potential pathogens, studies of HIV and SIV pathogenesis should take them into account as possible contributors to disease progression. This provides substantial opportunity to explain and eventually intervene in the processes that lead to AIDS clinical disease progression. Our data indicate that the expansion of the enteric virome can be used as a marker for rapidly progressive disease.

**[0089]** Accordingly, the present invention also relates to methods of detecting acquired immune deficiency syndrome (AIDS) and/or AIDS progression in a subject by detecting one or more target nucleotide sequences from a sample of the subject that specifically hybridize under stringent conditions to one or more of the polynucleotides of the invention, where the detection of an increase in the level of the one or more target nucleotide sequences in the subject, relative to the level of one or more target nucleotide sequences from a control subject, indicates AIDS and/or AIDS progression in the subject. Detecting of the one or more target nucleotide sequences may include synthesizing cDNA from RNA of the sample, and may utilize a PCR assay for detection, such as a real time PCR (RT-PCR) assay and/or a nested PCR assay. SEQ ID NOs: 332-371 (see, for example, Table 1) disclose primers that can be utilized in a PCR assay to screen for the presence of the viruses. The sample may be a tissue, organ, liquid, or feces sample from a mammal, preferably a primate or a human. This method of detecting AIDS and/or AIDS progression in a subject can be used alone, in conjunction, or in parallel with known method(s) of detecting AIDS and/or AIDS progression, such as by the detection of CD4 T cell levels.

#### Diagnosis of Immunodeficiency or Enteropathy

**[0090]** The compositions of the invention may be used for other diagnostic purposes. In some aspects, the invention features methods of diagnosing, or providing a prognostic indicator of, immunodeficiency and/or enteropathy in a subject including detecting one or more target nucleotide sequences from a sample of the subject that specifically hybridize under stringent conditions to one or more of the polynucleotides of the invention, where the detection of an increase in the level of the one or more target nucleotide sequences in the subject, relative to the level of one or more target nucleotide sequences from a control subject, indicates the presence of, or the propensity to develop, immunodeficiency and/or enteropathy in the subject. Detecting one or more target nucleotide sequences may include synthesizing cDNA from RNA of the sample, and may utilize a PCR assay for detection, such as a real time PCR (RT-PCR) assay and/or a nested PCR assay. SEQ ID NOs: 332-371 (see, for example, Table 1) disclose primers that can be utilized in a PCR assay to screen for the presence of the viruses. The sample may be a tissue, organ, liquid, or feces sample from a mammal, preferably a primate or a human. This method of diagnosing, or providing a prognostic indicator of, immunodeficiency and/or enteropathy in a subject can be used alone, in conjunction, or in parallel with known method(s) of diagnosing, or providing a prognostic indicator of, immunodeficiency and/or enteropathy, such as by the detection of CD4 T cell levels and/or serum LPS binding protein (LBP) levels.

#### Treatment of Immunodeficiency or Enteropathy

**[0091]** In other aspects, the compositions of the invention may be used for therapeutic purposes. For example, the invention features nucleic acid- or polypeptide-based vaccines. The vaccines may include a vector that includes the polynucleotides of the invention or a vaccine that includes a polypeptide of the invention. In addition, the invention features methods of identifying a therapeutic agent for use in treating immunodeficiency and/or enteropathy in a subject including detecting one or more target nucleotide sequences that specifically hybridize under stringent conditions to one or more of the polynucleotides of the invention from a sample of a subject administered a therapeutically effective amount of a candidate agent, where the detection of a decrease in the level of the one or more target nucleotide sequences in the subject, relative to the level of one or more target nucleotide sequences from the subject prior to administration or a control subject, identifies the candidate agent as the therapeutic agent. In some embodiments, the candidate agent is administered to the subject in a therapeutically effective amount. In other embodiments, the immunodeficiency and/or enteropathy is associated with a lentivirus (e.g., human immunodeficiency virus (HIV) or simian immunodeficiency virus (SIV)). Detecting one or more target nucleotide sequences may include synthesizing cDNA from RNA of the sample, and may utilize a PCR assay for detection, such as a real time PCR (RT-PCR) assay and/or a nested PCR assay. SEQ ID NOs: 332-371 (see, for example, Table 1) disclose primers that can be utilized in a PCR assay to screen for the presence of the viruses. The sample may be a tissue, organ, liquid, or feces sample from a mammal, preferably a primate or human.

#### Administration of a Therapeutic Agent

**[0092]** The vaccines of the invention or the therapeutic agent, once identified by the methods of the invention, can be administered to a subject (e.g., a human), pre- or post-lentiviral (e.g., HIV) infection, to treat, prevent, ameliorate, inhibit the progression of, or reduce the severity of immunocompromise and/or enteropathy. The subject, at the time of administration, may present as symptomatic or asymptomatic. In addition, the vaccine or identified therapeutic agent may also treat, prevent, ameliorate, inhibit the progression of, or reduce the severity of one or more symptoms, if present, of lentiviral (e.g., HIV) infection. Examples of the symptoms caused by lentiviral infection include one or more of, e.g., fever, muscle aches, coughing, sneezing, runny nose, sore throat, headache, chills, diarrhea, vomiting, rash, weakness, dizziness, bleeding under the skin, in internal organs, or from body orifices like the mouth, eyes, or ears, shock, nervous system malfunction, delirium, seizures, renal (kidney) failure, personality changes, neck stiffness, dehydration, seizures, lethargy, paralysis of the limbs, confusion, back pain, loss of sensation, impaired bladder and bowel function, and sleepiness that can progress into coma or death. These symptoms, and their resolution during treatment, may be measured by, e.g., a physician during a physical examination or by other tests and methods known in the art.

**[0093]** The vaccines or therapeutic agents can be formulated for administration alone or as a pharmaceutical composition by a route selected from, e.g., intramuscular, intravenous, intradermal, intraarterial, intraperitoneal, intralesional, intracranial, intraarticular, intraprostatic, intrapleural, intratracheal, intranasal, intravitreal, intravaginal, intrarectal,

topical, intratumoral, peritoneal, subcutaneous, subconjunctival, intravesicular, mucosal, intrapericardial, intraumbilical, intraocular, oral, or local administration, or by inhalation, by injection, by infusion, by continuous infusion, by localized perfusion bathing target cells directly, by catheter, by lavage, by gavage, in cremes, or in lipid compositions. The preferred method of administration can vary depending on various factors (e.g., the components of the composition being administered and the severity of the condition being treated). Formulations suitable for oral or nasal administration may consist of liquid solutions, such as an effective amount of the composition dissolved in a diluent (e.g., water, saline, or PEG-400), capsules, sachets, tablets, or gels, each containing a predetermined amount of the chimeric Ad5 vector composition of the invention. The pharmaceutical composition may also be an aerosol formulation for inhalation, e.g., to the bronchial passageways. Aerosol formulations may be mixed with pressurized, pharmaceutically acceptable propellants (e.g., dichlorodifluoromethane, propane, or nitrogen). In particular, administration by inhalation can be accomplished by using, e.g., an aerosol containing sorbitan trioleate or oleic acid, for example, together with trichlorofluoromethane, dichlorofluoromethane, dichlorotetrafluoroethane, or any other biologically compatible propellant gas.

**[0094]** Immunogenicity of the vaccine of the invention may be significantly improved if it is co-administered with an immunostimulatory agent or adjuvant. Suitable adjuvants well-known to those skilled in the art include, e.g., aluminum phosphate, aluminum hydroxide, QS21, Quil A (and derivatives and components thereof), calcium phosphate, calcium hydroxide, zinc hydroxide, glycolipid analogs, octodecyl esters of an amino acid, muramyl dipeptides, polyphosphazene, lipoproteins, ISCOM matrix, DC-Chol, DDA, cytokines, and other adjuvants and derivatives thereof.

**[0095]** The vaccine or therapeutic agent of the invention, or a pharmaceutical composition including the same, may be formulated to release the vaccine or therapeutic agent immediately upon administration (e.g., targeted delivery) or at any predetermined time period after administration using controlled or extended release formulations. Administration in controlled or extended release formulations is useful where the vaccine or agent, either alone or in combination, has (i) a narrow therapeutic index (e.g., the difference between the plasma concentration leading to harmful side effects or toxic reactions and the plasma concentration leading to a therapeutic effect is small; generally, the therapeutic index, TI, is defined as the ratio of median lethal dose ( $LD_{50}$ ) to median effective dose ( $ED_{50}$ )); (ii) a narrow absorption window at the site of release (e.g., the gastro-intestinal tract); or (iii) a short biological half-life, so that frequent dosing during a day is required in order to sustain a therapeutic level.

**[0096]** Many strategies can be pursued to obtain controlled or extended release in which the rate of release outweighs the rate of metabolism of the vaccine or therapeutic agent, or the pharmaceutical composition including the same. For example, controlled release can be obtained by the appropriate selection of formulation parameters and ingredients, including, e.g., appropriate controlled release compositions and coatings. Suitable formulations are known to those of skill in the art. Examples include single or multiple unit tablet or capsule compositions, oil solutions, suspensions, emulsions, microcapsules, microspheres, nanoparticles, patches, and liposomes.

**[0097]** The vaccine or therapeutic agent may be administered, e.g., 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 30, 35, 40, 45, 50, 55, or 60 minutes, 2, 4, 6, 10, 15, or 24 hours, 2, 3, 5, or 7 days, 2, 4, 6 or 8 weeks, or even 3, 4, or 6 months pre-exposure, or may be administered to the subject 15-30 minutes or 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 12, 20, 24, 48, or 72 hours, 2, 3, 5, or 7 days, 2, 4, 6 or 8 weeks, 3, 4, 6, or 9 months, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20 years or longer post-exposure to the infective agent, alone or in a pharmaceutical composition.

**[0098]** When treating disease (e.g., AIDS due to HIV infection), the vaccine or therapeutic agent may be administered to the subject either before a definitive diagnosis, before the occurrence of immunodeficiency and/or enteropathy, or after diagnosis or symptoms become evident. For example, the pharmaceutical composition including the vaccine or therapeutic agent may be administered, e.g., immediately after diagnosis or the clinical recognition of symptoms or 2, 4, 6, 10, 15, or 24 hours, 2, 3, 5, or 7 days, 2, 4, 6 or 8 weeks, or even 3, 4, or 6 months after diagnosis or detection of symptoms.

**[0099]** The vaccines or therapeutic agents may be sterilized by conventional sterilization techniques, or may be sterile filtered. The resulting aqueous solutions may be packaged for use as is, or lyophilized; the lyophilized preparation may be administered in powder form or combined with a sterile aqueous carrier prior to administration. The pH of the preparations typically will be between 3 and 11, more preferably between 5 and 9 or between 6 and 8, and most preferably between 7 and 8, such as 7 to 7.5. The resulting compositions in solid form may be packaged in multiple single dose units, each containing a fixed amount of the vaccine or therapeutic agent, and, if desired, one or more immunomodulatory agents, such as in a sealed package of tablets or capsules, or in a suitable dry powder inhaler (DPI) capable of administering one or more doses.

#### Dosages

**[0100]** The dose of the vaccine or therapeutic agent or the number of treatments using the same may be increased or decreased based on the severity of, occurrence of, or progression of, the level of immunocompromise and/or enteropathy in the subject (e.g., based on the severity of one or more symptoms of, e.g., viral infection). The dosage administered depends on the subject to be treated (e.g., the age, body weight, capacity of the immune system, and general health of the subject being treated), the form of administration (e.g., as a solid or liquid), the manner of administration (e.g., by injection, inhalation, dry powder propellant), and the cells targeted (e.g., epithelial cells, such as blood vessel epithelial cells, nasal epithelial cells, or pulmonary epithelial cells).

**[0101]** In addition, single or multiple administrations of the vaccines or therapeutic agents of the present invention may be given (pre- or post-infection) to a subject (e.g., one administration or administration two or more times). For example, subjects who are particularly susceptible to, e.g., viral infection may require multiple treatments to establish and/or maintain protection against the virus. Levels of induced immunity provided by the vaccines or therapeutic agents described herein can be monitored by, e.g., measuring CD4 T cell levels and/or serum LPS binding protein (LBP) levels. The dosages may then be adjusted or repeated as necessary to maintain desired therapeutic levels in the subject having immunocompromise and/or enteropathy associated with, e.g., a lentiviral (e.g., HIV) infection.

**[0102]** In some embodiments, a single dose of the vaccine or therapeutic agent may achieve protection, pre-exposure, from infective agents. In addition, a single dose administered post-exposure to a viral or other infective agent can function as a treatment according to the present invention. Multiple doses (e.g., 2, 3, 4, 5, or more doses) can also be administered, in necessary, to these subjects.

#### Carriers, Excipients, Diluents

**[0103]** The compositions of the invention may include a recombinant replication-defective Ad5 vector with chimeric hexon and fiber proteins, containing a heterologous nucleic acid molecule encoding an antigenic gene product or fragment thereof. An adenoviral vector of the invention also includes one or more of the adenoviruses identified in the present study (e.g., one or more of these adenoviruses may be used as a vector that is modified to include a heterologous nucleic acid molecule, which, upon expression in a host, produces a therapeutic immunogenic response in the host). Therapeutic formulations of the compositions of the invention are prepared using standard methods known in the art by mixing the active ingredient having the desired degree of purity with optional physiologically acceptable carriers, excipients or stabilizers (Remington's Pharmaceutical Sciences (20<sup>th</sup> edition), ed. A. Gennaro, 2000, Lippincott, Williams & Wilkins, Philadelphia, Pa.). Acceptable carriers, include saline, or buffers such as phosphate, citrate and other organic acids; antioxidants including ascorbic acid; low molecular weight (less than about 10 residues) polypeptides; proteins, such as serum albumin, gelatin or immunoglobulins; hydrophilic polymers such as polyvinylpyrrolidone, amino acids such as glycine, glutamine, asparagine, arginine or lysine; monosaccharides, disaccharides, and other carbohydrates including glucose, mannose, or dextrans; chelating agents such as EDTA; sugar alcohols such as mannitol or sorbitol; salt-forming counterions such as sodium; and/or nonionic surfactants such as TWEEN<sup>TM</sup>, PLURONICS<sup>TM</sup>, or PEG.

**[0104]** Optionally, the formulation contains a pharmaceutically acceptable salt, preferably sodium chloride, e.g., at about physiological concentrations. Optionally, the formulations of the invention can contain a pharmaceutically acceptable preservative. In some embodiments the preservative concentration ranges from 0.1 to 2.0%, typically vv. Suitable preservatives include those known in the pharmaceutical arts. Benzyl alcohol, phenol, m-cresol, methylparaben, and propylparaben are examples of preservatives. Optionally, the formulations of the invention can include a pharmaceutically acceptable surfactant at a concentration of 0.005 to 0.02%.

**[0105]** These and other aspects of the invention are further described in the Examples, below.

#### EXAMPLES

**[0106]** The following examples are to illustrate the invention. They are not meant to limit the invention in any way.

##### Example 1

##### Materials and Methods

##### Nucleic Acid Preparation and 454 Sequencing

**[0107]** 100 mg or 200 mg of frozen stool was resuspended in 6 volumes of PBS (Finkbeiner et al., *PLoS Pathog.* 4:

e1000011 (2008)), centrifuged to pellet particulate matter and the supernatant was then passed through a 0.45- $\mu$ m filter. Total nucleic acid was isolated from 200  $\mu$ L or 850  $\mu$ L of this filtrate using the Ampliprep DNA extraction machine (Roche) according to manufacturer's instructions. To enable detection of both RNA and DNA viruses, 9  $\mu$ L total nucleic acid from each sample was reverse transcribed and 6  $\mu$ L of the cDNA reaction amplified as previously described (Wang et al., *PLoS Biol.* 1: E2 (2003)). Briefly, RNA templates were reverse transcribed using a first primer containing a 16-nucleotide specific sequence followed by 9 random nucleotides for random priming. The 16-nucleotide specific sequence was unique for each sample and served as a barcode in assigning sequencing sequences to a sample. Sequenase (United States Biochemical) was used for second strand cDNA synthesis and for random-primed amplification of DNA templates using the first primer. Each sample was subjected to 40 cycles of PCR amplification using a second primer containing the same 16 nucleotide specific sequence as in the corresponding first primer. Amplification products were quantitated, diluted to 15 ng/ $\mu$ L and then 5  $\mu$ L of each sample was pooled, adaptor-ligated and sequenced on the 454 GS-FLX platform (454 Life Sciences).

#### Detection and Analysis of Viral Sequences Using Custom Bioinformatic Pipeline

**[0108]** Sequences were analyzed using VirusHunter as described (Presti et al., *J. Virol.* 83: 11599-11606 (2009); Loh et al., *J. Virol.* 83: 13019-13025 (2009); Zhao et al., *J. Virol.* 85: 10230-10238 (2011); Felix et al., *PLoS Biol.* 9: e1000586 (2011); Loh et al., *J. Virol.* 85: 2642-2656 (2011)). Briefly, sequences were assigned to samples based on the unique barcode sequences (i.e., the second primer sequences), primer sequences were trimmed, and sequences were clustered using CD-HIT (Li et al., *Bioinformatics* 22: 1658-1659 (2006)) to remove redundant sequences (95% identity over 95% sequence length). The longest sequence from each cluster was chosen as the representative unique sequence and entered into the analysis pipeline. Then, unique sequences were masked by RepeatMasker (Smit, et al. *RepeatMasker Open-3.0*). If a sequence did not contain a stretch of at least 50 consecutive non-"N" nucleotides or if greater than 40% of the total length of the sequence was masked, it was removed from further analysis (filtered). Filtered high quality unique non-repetitive sequences were sequentially compared against (i) the human genome using BLASTn; (ii) GenBank nt database using BLASTn; and (iii) GenBank nr database using BLASTX (Altschul et al., *J. Mol. Biol.* 215: 403-410 (1990)). Minimal e-value cutoffs of  $1e^{-10}$  and  $1e^{-5}$  were applied for BLASTn and BLASTX, respectively (Bench et al., *Appl. Envir. Microbiol.* 73: 7629-7641 (2007); Wommack et al., *Appl. Envir. Microbiol.* 74: 1453-1463 (2008)). Sequences were phylotyped as human, mouse, fungal, bacterial, phage, viral, or other based on the identity of the top BLAST hit. Sequences without any significant hit to any of the databases were designated as unassigned. If a sequence aligned to both a virus and another kingdom (e.g., bacteria or fungi) with the same e-value it was classified as ambiguous. All eukaryotic viral sequences were further classified into viral families based on the taxonomy ID of the best hit.

#### Assembly of Viral Contigs and Virus Comparison Analysis

**[0109]** All viral sequences, unassigned sequences, and the longest five similar sequences for those sequences from each

sample were assembled into contigs using Newbler (454 Life Sciences) with default parameters. If a sample was sequenced multiple times, all available sequencing data were used to optimize contig assembly. The longest contig from amongst all contigs belonging to a given genus was chosen as the first representative contig. To compare viruses across multiple animals all sequences (contigs and sequences if no contigs were obtained from a sample) were compared with this representative virus contig. If a sequence aligned with the representative contig over its full length and shared 98% nucleotide identity or higher over the aligned region it was considered to be the same as the representative contig. For sequence that was considered as different from the representative contig, the next longest contig was selected as the second representative virus. This process was repeated until all sequences were classified. If two contigs or sequences were located at different regions of the genome, and no conclusive decision could be made about their possible relatedness, we defaulted in a conservative fashion to assuming that only a single virus was present. Representative viral contigs were queried against the NCBI nt database and the most related viral genomes were identified. The most closely related virus with full genome sequence available was selected as the reference genome. For adenoviruses different sequences shared the highest homology with different viruses, indicating that in these large genomes some regions of the new viruses we detected were most related to different viruses in the data base. Two out of the three contig sequences used for designing primers shared highest homology to simian adenovirus 1 strain ATCC VR-195, which was therefore selected as reference genome. If no nucleotide level homology was detected, viral contigs were queried for protein homology against the NCBI nr database and the most related viral genome was identified.

#### Metagenomic Analysis Using MEGAN

**[0110]** Individual sequences obtained by 454 sequencing were analyzed using BLASTX (version 2.2.22+) on a customized server with ~1700 available processor slots and a memory range of 2-32 GB per node. Sequences were compared by BLASTX to the NCBI nr database version Jun. 6, 2011. Results with an e-value  $\leq 1e^{-10}$  were stored and used for taxonomic assignment using the Lowest-Common Ancestor (LCA) algorithm in MEGAN v. 4.62.3 (22 Nov. 2011). The following LCA parameters were used for taxonomic assignment: Min Support: 5, Min Score: 35, Top percent: 10, Win Score: 0, Min Complexity: 0. This process resulted in the generation of sample specific RMA files used by MEGAN for downstream analysis. These files contain all of the taxonomic assignment information for each sample. Global metagenome comparisons using all sequences assigned to all taxa were completed for each cohort. These comparisons used MEGAN's normalization protocol enabling inter-sample comparison. Additionally, sequences contained in specific taxonomic subsets (bacteria, viruses, or phage) were isolated and processed through MEGAN using the same parameters. Similarly, sequences from specific phage taxa (caudovirales, microviridae, leviviridae and unclassified phage) were extracted and compared. This procedure permitted independent analysis of these taxa without artifactual effects of global normalization. Summarized sequence counts per taxa were exported for subsequent statistical analysis using Graph Pad Prism version 5.0d.

## PCR Detection of Viruses

[0111] Primers (Table 1) were designed to amplify regions conserved between WUHARV adenoviruses 1-5, caliciviruses 1-2, calicivirus 3, parvoviruses 1-2, enteroviruses 1-3, sapeloviruses 1-3, and related viral genomes. Primer sensitivity was evaluated using libraries with high or low numbers of adenovirus, calicivirus, parvovirus, enterovirus, or sapelovirus sequences, while primer specificity was evaluated using libraries with high numbers of unrelated virus

sequences, as well as virus sequences from related genera. Libraries generated from stool samples were diluted 10 fold and screened (n=2) for presence of viruses using: 10×PCR buffer 2.5 ul, MgCl<sub>2</sub> (25 mM) 2.5 ul, dNTP (2 mM) 2.5 ul, forward primer (10 uM) 0.5 ul, reverse primer (10 uM) 0.5 ul, Taq 0.3 ul, and H<sub>2</sub>O 6.2 ul. PCR products were amplified at: 95° C., 5 min; 95° C., 30 sec, 60° C., 30 sec, 72° C., 1 min, for 32 cycles; 72° C., 10 min and then visualized using EtBr on a 1.5% agarose gel. There was concordance in all duplicate tests.

TABLE 1

Primers					
WUHARV Virus	Primer name	Targeted region	Sequence (5'-3')	SEQ ID NO:	Orientation
Adenovirus 1	4302c3f	Hexon	GGCAATCATGATGGACACCT T	332	F
Adenovirus 1	4302c3r	Hexon	TTAATCACCACCGCAACGC	333	R
Adenovirus 1	4302c1f	Hexon	CAATGGAACATTAATCCCAC G	334	F
Adenovirus 1	4302c1r	Hexon	CCTGCCAACACTCCCATATT T	335	R
Adenovirus 1	4302c18f	E1B	AGAGCTATCACACAGCGTTC A	336	F
Adenovirus 1	4302c18r	E1B	ACCGAGTGGTGGAGGAGAA	337	R
Adenovirus 2	4310ac18f	pIIIa	TAACGTTCAGACCAATCTGG A	338	F
Adenovirus 2	4310ac18r	pIIIa	CGGCAATAGTGCTACTGTTG G	339	R
Adenovirus 2	4310ac16f	Hexon	CGGGACAACCTTCATTGGACT	340	F
Adenovirus 2	4310ac16r	Hexon	GCGCCAATGTTTACAAAGGT	341	R
Adenovirus 3	4310bc18f	pIIIa	TAACGTTCAGACCAATCTGG A	342	F
Adenovirus 3	4310bc18r	pIIIa	CGGCAATAGTGCTACTGTTG G	343	R
Adenovirus 3	4310bc21f	Hexon	ACGACAGCACCAGTTCAAAA C	344	F
Adenovirus 3	4310bc21f	Hexon	TTTTCTGGCAGCGTGATGTT	345	R
Adenovirus 3	4310bc28r	E3	CTCTTGGCAACCCCTTATTG	346	F
Adenovirus 3	4310bc28f	E3	TGGGTGAAACCATTCTGTGTT	347	R
Adenovirus 4	4312u11r	E3	CCGTCCTCTCTGGTAGAAA	348	F
Adenovirus 4	4312u11f	E3	CGTCGACTGTTGGAGAAACA	349	R
Adenovirus 4	4312u10r	DBP	GCCGTTACATCCAGATCCTC	350	F
Adenovirus 4	4312u10f	DBP	TACACGAGGGAATGAAAGC	351	R

TABLE 1-continued

Primers					
WUHARV Virus	Primer name	Targeted region	Sequence (5'-3')	SEQ ID NO:	Orientation
Adenovirus 4	4312u7r	NCR <sup>1</sup> between E1a and E1b	CTTGTGCCTGTGCTTTTCAT	352	F
Adenovirus 4	4312u7f	NCR between E1a and E1b	GTGCAAAGAGAACTAGTATG G	353	R
Adenovirus 5	4287u7f	Iva	GGATGTTCAAGTACATGGGC A	354	F
Adenovirus 5	4287u7r	Iva	GATGCATGACAAGTTCCCA A	355	R
Adenovirus 5	4287c5f	E3	GAATGGTAGCTGCTTTCTTC A	356	F
Adenovirus 5	4287c5r	E3	TGTTGGGTGATTGTGATGGA	357	R
Adenovirus 5	4287c11f	Fiber-1	CTGAAAAAACGAATTGGTG G	358	F
Adenovirus 5	4287c11r	Fiber-1	TTGACAACAATGGTGCCTG	359	R
Adenovirus (1-5)	Adv-a GG	pIIIa	ACTAACGTKCAGACCAATCT	360	F
Adenovirus (1-5)	Adv-b	pIIIa	GTACAGRCTCACGGACTGC	361	R
Calicivirus (1-2)	CV-a	NS <sup>2</sup> polyprotein	GTACGAYGTCGGAGGGACC	362	F
Calicivirus (1-2)	CV-b	NS polyprotein	GRTCACAAGCCATGACACTC AG	363	R
Calicivirus 3	CV-c	NS polyprotein	TTATGTTATGGACAACCCAA AGG	364	F
Calicivirus 3	CV-d	NS polyprotein	GGTCAAGAGACAATAGCTCC AT	365	R
Parvovirus (1-3)	PV-a	capsid	ACCAGACTAACWCAAGGCG C	366	F
Parvovirus (1-3)	PV-b	capsid	GGTASGTGTTCCATTGTCTT GG	367	R
Enterovirus (1-3)	EV-a	5' UTR <sup>3</sup>	GCACAACCCAGTGTAGTTC	368	F
Enterovirus (1-3)	EV-b	5' UTR	CCAATCCAATMGCTATATGA TGAC	369	R
Sapelovirus (1-3)	SV-a	5' UTR	CCAGKMTAAAAGGCAATTGT GG	370	F
Sapelovirus (1-3)	SV-b	5' UTR	CCTGTCAGGTAGCACTAGAC T	371	R

<sup>1</sup>NCR = non-coding region<sup>2</sup>NS = non-structural<sup>3</sup>UTR = untranslated region

## Isolation and Detection of WUHARV Adenoviruses

**[0112]** Stool samples from rhesus monkeys #30, 40, and 44 were diluted in media, passed through a 0.45- $\mu$ m filter, and used to inoculate a T-25 flask containing an E1 complementing cell line such as PerC6 or Per55K cells maintained as previously described (Abbink et al., J. Virol. 81: 4654-4663 (2007)). Upon 100% cell lysis, cells and supernatant were

harvested and frozen at  $-20^{\circ}$  C. Viruses were plaque purified twice. Briefly, MW6 plates were seeded with Per55K cells on day 1. On day 0 cells were infected with log dilutions of virus. On day 1 an agar overlay was performed, plates were incubated until plaques were big enough to pick, and plaques picked and amplified in a well of a 24 well plate. Virus stocks were then generated and virions purified. Briefly, virus was

amplified to inoculate 24 T-175 triple layer flasks. Cells were harvested and virus particles purified using CsCl. To detect adenoviruses, primers (Supplementary Table 1) were designed to amplify regions from WUHARV adenoviruses (1-5) from contigs with a range of relatedness to the reference genomes. Crude lysate, plaques and purified virus were screened for presence of adenovirus using: 2 ul DNA, 25 ul Phusion Master Mix with HF buffer, 1.5 ul 100% DMSO, 2 ul forward primer (10 uM), 2 ul reverse primer (10 uM), 17.5 ul H<sub>2</sub>O. PCR products were amplified at: 98° C. 30 sec; 98° C. 10 sec, 50° C. 10 sec, 72° C. 30 sec for 30 cycles; 72° C. 10 min. and then visualized using EtBr on a 0.8% agarose gel.

#### Assays and Necropsy of SIV-Infected Rhesus Monkeys

**[0113]** Serum levels of LPS binding protein (LBP) were quantitated by ELISA (Antibodies Online). Twelve animals housed at the NEPRC were subjected to complete necropsy within two hours of death and representative sections of all major organs were collected, fixed in 10% neutral buffered formalin (NBF), embedded in paraffin, sectioned at 5 µm, and stained using haematoxylin and eosin (HE). Following histopathologic examination, additional immunohistochemistry was used to analyze the degree of adenovirus infection within the small and large intestinal sections. The specific adenovirus immunohistochemistry protocol was as follows: deparaffinization and rehydration followed by a 5' block in 3% hydrogen peroxide; pre-treatment with proteinase K for 5 minutes; all steps were followed by a tris-buffered saline (TBS) wash. Prior to application of primary antibodies, all slides were treated with both a biotin block and a Dako protein block for 10 minutes each. Sections were incubated with anti-mouse adenovirus known to cross react with 41 known serovars of adenovirus (Millipore (Billerica, Mass., USA), monoclonal, 1:200) overnight at 4° C. This was followed by 30 minute incubation at room temperature with Vectastain ABC standard. All slides were developed with DAB chromagen (Dako) and counterstained with Mayer's haematoxylin. In all cases, step sections were incubated with isotype-specific irrelevant antibodies for negative controls and failed to show staining. Positive controls consisted of sections of small intestine positive for adenovirus.

#### GenBank Accession Numbers

**[0114]** Sequence data from each animal were uploaded to the MG-RAST server (version 3.12). The sequences of viral contigs presented in FIG. 5 have been uploaded to GenBank with the following numbers: WUHARV Calicivirus 1 (JX627575), WUHARV Parvovirus 1 (JX627576), WUHARV Enterovirus 1 (JX627570), WUHARV Enterovirus 2 (JX627571), WUHARV Enterovirus 3 (JX627572), WUHARV Sapelovirus 1 (JX627573), and WUHARV Sapelovirus 2 (JX627574).

#### Statistical Analysis

**[0115]** For analysis of sequence numbers after normalization the data were log<sub>10</sub> transformed prior to statistical analysis. P-values were derived using the nonparametric Mann-Whitney test. P-values <0.05 are considered significant. For analysis of bacterial families in FIG. 9, we utilized one-way ANOVA with a Bonferroni correction to correct for multiple comparisons.

#### Construction of Phylogenetic Trees

**[0116]** We performed phylogenetic analysis for viruses with sufficient sequence information as defined by contig length is >90% of full length of the most closely related viruses shown in FIG. 5. Multiple sequence alignments were performed with ClustalW (Thompson et al. *Nucleic Acids Res.* 22: 4673-4680 (1994)). Phylogenetic analysis was performed using the neighbor-joining method in the PHYLIP package (Felsenstein, *Phylogeny Inference Package*, Department of Genome Sciences, University of Washington, Seattle (2005)) with 100 bootstrap replicates. Phylogenetic trees were visualized using TreeView (Page, *CABIOS*. 12: 357-358 (1996)).

#### Caliciviridae Sequences Used for Phylogenetic Trees

**[0117]** The predicted amino acid sequences of the full length polyprotein from WUHARV Calicivirus 1 were used to construct a phylogenetic tree. Polyproteins from following viruses were used: Bovine calicivirus (BoCAA09480.1), Calicivirus pig/F15-10/CAN (CV pig F15-10, ACQ44561.1), Calicivirus pig/AB104/CAN (CV pig AB104, ACQ44563.1), Calicivirus pig/NC-WGP93C/USA/2009 (CV pig NC-WGP93C, ADG27878.1), Calicivirus pig/AB90/CAN (CV pig AB90, YP\_002905325.1), Norovirus Hu/GII-4/Niigata2/2008/JP (BAJ13866.1), Norovirus dog/GVI.1/HKU\_Ca026F/2007/HKG (ACV89839.1), Norovirus genogroup 3 (AFQ00092.1), Norovirus Bo/Newbury2/1976/UK (AAD16174.5), Norwalk-like virus (AAM95184.2), Norwalk virus (NP\_056820.1), Tulane Virus (ACB38131.1), and WUHARV Calicivirus 1 (JX627575).

#### Parvoviridae Sequences Used for Phylogenetic Trees

**[0118]** The predicted amino acid sequences of the near full length nonstructural 1 protein from WUHARV Parvovirus 1 were used to construct the phylogenetic tree. Polyproteins from following viruses were used: Bufavirus 1 (AFN44273.1), Bufavirus2 (AFN44276.1), Canine parvovirus (CPV\_AEK69509, AEK69509.1), Canine parvovirus (CPV\_AAV54174, AAV54174.1), Feline panleukopenia virus (FPV\_BAA 19018, BAA 19018.1), Feline panleukopenia virus (FPV\_AAC37927, AAC37927.1), Kilham rat virus (AAC40695.1), LuIII virus (NP\_821154.1), Mink enteritis virus (AE092090.1), Minute virus of mice (ABB01353.1), Mouse parvovirus 1 (NP\_042345.1), Mouse parvovirus 2 (YP\_656490.1), Parvovirus H1 (NP\_040318.1), Porcine parvovirus (ADN94624.1), Porcine parvovirus (ADN94588.1), and WUHARV Parvovirus 1 (JX627576).

#### Picornaviridae Sequences Used for Phylogenetic Trees

**[0119]** The full length genome of WUHARV Enterovirus 1, 2, 3, WUHARV Sapelovirus 1 and 2 were used to construct the phylogenetic tree. Genome sequences of following viruses were used: Baboon enterovirus strain A13 (BaboonEV A13, AF326750.2), Duck picornavirus TW90A (AY563023.1), Enterovirus 75 strain USA/OK85-10362 (EV 75, AY556070.1), Human echovirus 11, isolate HUN-1108 (HEchoV 11, AJ577589.1), Human enterovirus 71 strain BJ08-Z025-5 (HEV 71, FJ606450.1), Human enterovirus 90 (HEV 90, AB192877.1), Human enterovirus 92 strain RJG7 (HEV 92, EF667344.1), Human coxsackievirus A2 strain CVA2/SD/CHN/09 (HCoxV A2, HQ728259.1), Human coxsackievirus A5 strain CVA5/SD/CHN/09 (HCoxV A5,

HQ728261.1), Human coxsackievirus A7 strain Parker (HCox A7, AY421765.1), Porcine enterovirus 8 strain V13 (PSV-1, Porcine sapelovirus 1, AF406813.1), Porcine sapelovirus strain csh (PSV\_csh, HQ875059.1), Simian enterovirus 46 strain RNM5 (SimianEV 46, EF667343.1), Simian enterovirus SV19 strain M19s (SV19, AF326754.2), Simian enterovirus SV43 strain OM112t (SV43, AF326761.2), Simian sapelovirus 1 strain 2383 (SimianSV-1, AY064708.1), WUHARV Enterovirus 1 (JX627570), WUHARV Enterovirus 2 (JX627571), WUHARV Enterovirus 3 (JX627572), WUHARV Sapelovirus 1 (JX627573), and WUHARV Sapelovirus 2 (JX627574).

### Example 2

#### Next Generation Sequencing Analysis Reveals Expansion of the Enteric Virome During Pathogenic SIV Infection

##### Defining the Enteric Virome

**[0120]** To define the effects of pathogenic and non-pathogenic SIV infection on the enteric virome, we shotgun sequenced libraries of fecal RNA+DNA from four independent cohorts of monkeys, each comprising SIV-infected and uninfected control animals. Two cohorts of pathogenically SIV-infected and uninfected control rhesus monkeys were housed at the New England Primate Research Center (NEPRC) or the Tulane National Primate Research Center (TNPRC) (Table 2). As expected, the set point of SIV in the

serum correlated with rapid progression to AIDS and death. The NEPRC cohort was sampled at both 24 and 64 weeks after SIV infection. Two cohorts of non-pathogenically SIV-infected and uninfected control African green monkeys were housed at the National Institutes of Health (NIH, vervet monkeys) or the NEPRC (sabaues monkeys) (Table 2).

**[0121]** Total RNA+DNA from fecal material were sequenced using 454 technology to leverage the resulting long sequences for robust assessment of taxonomy and assembly of viral genomes (Table 2). There was no statistical correlation between SIV infection status and either the number of total or unique sequences (viral plus other) obtained within any of the four cohorts. For each cohort, sequences were analyzed by two different computational approaches. In the first method, the taxonomic structure of the sequences was analyzed using MEGAN version 4.62.3 (build Nov. 22, 2011 (Huson et al., *Genome. Res.* 17: 377-386 (2007); Huson et al., *BMC Bioinformatics* 10(Suppl 1): S12 (2009))). Each sequence was compared to the non-redundant (nr) database using BLASTX and results mapped to the NCBI Taxonomy Database. Sequences assigned to bacterial families or classes were extracted and used for subsequent analysis. The second computational approach was a custom pipeline called VirusHunter developed to identify novel viruses via analysis of both nucleic acid and protein similarity (Presti et al., *J. Virol.* 83: 11599-11606 (2009); Loh et al., *J. Virol.* 83: 13019-13025 (2009); Zhao et al., *J. Virol.* 85: 10230-10238 (2011); Felix et al., *PLoS Biol.* 9: e1000586 (2011); Loh et al., *J. Virol.* 85: 2642-2656 (2011)).

TABLE 2

Cohorts and sequences analyzed							
Animal cohort	Type of monkey	Animal numbers		Total sequences	Unique sequences	Unique Sequences per sample	Unique sequences per sample
				(average length)	(average length)		
NEPRC <sup>1</sup> (24 wpi <sup>2</sup> )	Rhesus	22	22	899,947	356,521	4,689-	594-
	Control		SIV+	(358 bp)	(357 bp)	51,870	26,838
NEPRC (64 wpi)	Rhesus	22	12	705,429	263,430	6,132-	1,080-
	Control		SIV+	(341 bp)	(345 bp)	59,847	33,982
TNPRC <sup>3</sup>	Rhesus	29	13	1,409,046	557,518	9,188-	3,666-
	Control		SIV+	(296 bp)	(294 bp)	89,974	33,613
NIH <sup>4</sup>	African green	19	19	1,382,171	425,524	3,259-	1,382-
	Control		SIV+	(300 bp)	(301 bp)	127,567	33,464
NEPRC	African green	6	10	612,612	187,807	8,287-	2,118-
	Control		SIV+	(293 bp)	(279 bp)	194,880	55,158

<sup>1</sup>New England Primate Research Center

<sup>2</sup>wpi = weeks post-infection with SIV

<sup>3</sup>Tulane National Primate Research Center

<sup>4</sup>National Institutes of Health



#### Enteric Virome of Rhesus Monkeys Housed at the NEPRC

**[0122]** We first analyzed the enteric virome of 44 rhesus monkeys housed at the NEPRC comprised of 22 monkeys infected intrarectally with pathogenic SIVmac251 and 22 SIV-uninfected monkeys (herein termed controls) (FIGS. 1A, 1B, 2A, and 2B). Per standard husbandry procedures, SIV-infected and control rhesus monkeys were fed the same diet but housed separately. Analysis of this cohort confirmed SIV viremia in infected animals and revealed the expected decreases in CD4 T cell counts and increases in serum LBP levels consistent with intestinal leakage and consequent systemic immune activation at both 24 and 64 weeks after infection (FIGS. 3A-3C). We collected fecal specimens either at 24 (FIGS. 1A and 2A) or 64 weeks (FIGS. 1B and 2B) after SIV infection. Between these two collection times 10 SIV-infected rhesus monkeys were euthanized for progressive AIDS. As expected, the set point level of SIV in the serum of rhesus monkeys correlated with rapid progression to AIDS and death. No control animals died.

**[0123]** SIV infection was associated with a greater than 10-fold increase in the number of sequences from viruses ( $p < 0.0001$ ) and a decrease in sequences from bacteria ( $p = 0.003$ ) at 24 weeks post-infection (FIGS. 1A and 2A). At this time after SIV infection, there were no statistically significant SIV-associated changes in the total number of sequences from phages, alveolata (representing protists), viridiplantae (representing food sequences from plants), or other kingdoms and phyla (FIGS. 1A and 2A). Samples collected 40 weeks later (64 weeks after SIV infection) revealed increases in viral sequences in most of the surviving animals that showed low numbers of viral sequences 24 weeks after SIV infection (e.g., compare animals 23, 31, and 33 between FIGS. 1A and 1B). Differences between SIV-infected and control monkeys, similar to those observed at 24 weeks after SIV infection, were observed for both viral ( $p < 0.0001$ ) and bacterial ( $p = 0.035$ ) sequences at 64 weeks after infection (FIGS. 1B and 2B). By 64 weeks after SIV infection, the surviving SIV-infected monkeys showed significant decreases in the number of phage ( $p = 0.0320$ ), alveolata ( $p = 0.0183$ ), and viridiplantae ( $p = 0.0013$ ) sequences compared to controls (FIG. 2B). These data suggest that pathogenic SIV infection is associated with significant expansion in the enteric virome.

#### Enteric Virome of Rhesus Monkeys Housed at the TNPRC

**[0124]** To confirm our findings in pathogenically SIV-infected rhesus monkeys housed at the NEPRC, we analyzed an independent cohort of 13 rhesus monkeys infected intravaginally with SIVmac251 and 29 control rhesus monkeys housed at the TNPRC (Table 2; FIGS. 1C and 2C). SIV infection at the TNPRC was associated with a significant increase in viral sequences ( $p = 0.0420$ ) and decrease in bacterial sequences (0.0019). In the TNPRC cohort, the SIV-infected monkeys showed significant increases in the number of phage ( $p = 0.0133$ ) sequences (FIGS. 1C and 2C). Similar to the 24 week time point in the NEPRC cohort there were no significant changes in sequences from phage, alveolata, viridiplantae, or sequences from other kingdoms and phyla (FIGS. 1C and 2C). These results confirm that an expansion of the enteric virome is associated with pathogenic SIV infection in two independent cohorts of rhesus monkeys.

#### Enteric Virome of African Green Monkeys

**[0125]** We next assessed whether the pathogenic SIV infection-associated changes in the enteric virome observed in

rhesus monkeys (FIGS. 1A-1C and 2A-2C) were also seen during non-pathogenic SIV infection in African green monkeys (Table 2; FIGS. 1D-1E and 2D-2E). The vervet African green monkey cohort housed at the NIH (FIGS. 1D and 2D) was comprised of six monkeys infected intravenously with SIVagm90, two monkeys infected intravenously with SIVagmVer1, 11 monkeys naturally infected with SIV, and 19 uninfected control animals. The cohort of sabaeus African green monkeys housed at the NEPRC (Table 2; FIGS. 1E and 2E) was comprised of two monkeys infected intravenously with SIVagmMJ8, 8 monkeys infected intravenously with SIVagm9315BR and 6 uninfected control animals. Analysis of these two sets of sequences revealed an increase in phage sequences in the NIH cohort ( $p = 0.0331$ ) that was not observed in the NEPRC cohort, but no other significant SIV infection-associated changes were observed in either cohort including for the virome (FIGS. 1D-1E and 2D-2E). These data indicate that the expansion of the enteric virome observed during pathogenic SIV infection is not observed during non-pathogenic SIV infection. Importantly, these African green monkeys had been infected with non-pathogenic SIV for a prolonged period (a minimum of 3 years for the NIH cohort, and from 27 weeks (2 animals) to 2.6 years (8 animals) for the NEPRC cohort). Therefore, the lack of an increase in viral sequences in these SIV-infected animals is not attributable to them being infected for a shorter time than the pathogenically SIV-infected rhesus monkeys analyzed above.

#### Example 3

##### Viruses Present in SIV-Infected Rhesus and African Green Monkeys

**[0126]** We next defined the nature of the viral sequences that we detected in SIV-infected and uninfected monkeys using VirusHunter software (Presti et al., *J. Virol.* 83: 11599-11606 (2009); Loh et al., *J. Virol.* 83: 13019-13025 (2009); Zhao et al., *J. Virol.* 85: 10230-10238 (2011); Felix et al., *PLoS Biol.* 9: e1000586 (2011); Loh et al., *J. Virol.* 85: 2642-2656 (2011)). When a nucleotide sequence did not have significant similarity to the genome of an already sequenced virus, we analyzed the predicted translation products and selected the most closely related virus in the database for comparison. This analysis allowed us to determine which types of viruses were detected in individual animals in each cohort (FIGS. 4A-4E). Using conservative criteria we detected at least 32 distinct and previously undescribed viruses in the sequences generated from individual rhesus monkeys housed at the NEPRC alone (FIGS. 4A and 4B). Certain viruses were found in multiple different animals, indicating shared exposure to enteric viruses. We did not count circoviruses in this estimate due to their ubiquity and diversity. Importantly, we found no significant differences in known insect (Dicistroviridae, Iflaviridae) or plant viruses, which are presumably derived from the diet, comparing SIV-infected animals and control animals in any cohort (FIGS. 4A-4E). The lack of differences in viruses from insects and plants between SIV-infected and SIV uninfected monkeys provides an important internal control indicating that the process of library construction and analysis does not artificially expand the number of mammalian viral sequences in samples from SIV-infected rhesus monkeys.

**[0127]** Newly identified viruses included five adenoviruses, three caliciviruses, one papillomavirus, seven members

of the Parvoviridae (2 parvovirusamdoviruses, five dependoviruses, and one bocavirus), seven picobimaviruses, and seven members of the Picornavirales (three enteroviruses, 3 sapeloviruses, and one picornavirus), and one polyomavirus (FIGS. 4A and 4B; Table 3). Importantly, many SIV-infected rhesus monkeys at both the NEPRC and the TNPRC were shedding multiple potentially pathogenic viruses (FIGS. 4A-4C). The presence of multiple novel viruses, and of individual animals infected with multiple distinct viruses, was not regularly observed in control rhesus monkeys housed at the same locations. In striking contrast, cohorts of African green monkeys housed at either the NEPRC or the NIH were relatively free of virus infection whether SIV-infected or not (FIGS. 4D and 4E).

[0128] As previously observed by others using classical virologic methods (Wang et al., *J. Med. Primatol.* 36: 101-

107 (2007); Oberste et al., *J. Gen. Virol.* 88: 3360-3372 (2007); Oberste et al., *J. Virol.* 76: 1244-1251 (2002); Sasseville et al., *J. Immunotoxicol.* 7: 79-92 (2010); Bailey et al., *Vet. Pathol.* 47: 462-481 (2010)), picornaviruses were detected in both control and SIV-infected rhesus monkeys (FIG. 4; Table 3). This allowed us to compare the number of sequences detected from pathogenic SIV-infected rhesus monkeys versus control animals (FIG. 4F). In monkeys housed at either the NEPRC or the TNPRC there were significant increases in the number of sequences derived from picornaviruses in SIV-infected animals compared to controls ( $p=0.0002$  and  $0.0004$  for the NEPRC rhesus animals at 24 or 64 weeks of infection,  $p=0.0247$  for the TNPRC rhesus animals). No relationship was detected between picornavirus sequences and SIV with non-pathogenic SIV infection of African green monkeys. These data are consistent with a failure to control picornavirus infection in association with pathogenic SIV infection.

TABLE 3

Summary of viruses identified in Rhesus macaques at the NEPRC							
Virus family	Virus name <sup>1</sup>	Name of most closely related virus <sup>2</sup>	Percent identity nt or aa <sup>3</sup> (length, nt or aa)	Animal(s) in which viruses were detected <sup>4</sup>			
				24 weeks		64 weeks	
				SIV-	SIV+	SIV-	SIV+
Adenoviridae	WUHARV Adenovirus 1	Simian adenovirus 1 ATCC VR-195	79-99% nt	—	<u>40</u> <sup>5*</sup>	—	23
	WUHARV Adenovirus 2	Human adenovirus G	36-100% nt	—	<u>44</u> *	—	—
	WUHARV Adenovirus 3	Human adenovirus G	36-100% nt	—	<u>44</u> *	—	—
	WUHARV Adenovirus 4	Human adenovirus G	87-93% nt	—	<u>30</u>	—	—
	WUHARV Adenovirus 5	Human adenovirus G	48-100% nt	—	—	—	27
Caliciviridae	WUHARV Calicivirus 1	Tulane virus	75% nt (4839/6489)	—	25, 26, 31, 33, 34*, 35*, 37, 38, 39*, 40*, 41, 44*	—	23, 25
	WUHARV Calicivirus 2	Tulane virus	88-93% nt (4463/5082, 753/812)	—	37	—	30, 37, 41
	WUHARV Calicivirus 3	Rhesus macaque recovirus strain FT437	0-81% nt (50/64, 2881413, 216/268)	—	23, 28*, 32, 39*	—	—
Papillomaviridae	WUHARV Papillomavirus 1	Human papillomavirus	69% nt (300/432)	15	25	—	32
Parvoviridae	WUHARV Bocavirus 1	Human bocavirus isolate KU3	73% nt (118/160)	—	—	—	<u>29</u>
	WUHARV Dependovirus 1	Adeno-associated virus 11	93% nt (3812/4090)	—	<u>40</u> *	—	29
	WUHARV Dependovirus 2	Adeno-associated virus 10	92% nt (3680/4020)	—	23, 30, 31, 32, 34*, 38, 39*, 40*, 44*	—	27, 29
	WUHARV Dependovirus 3	Adeno-associated virus isolate rh.31	94% nt (1680/1793)	—	<u>40</u> *	—	—

TABLE 3-continued

Summary of viruses identified in Rhesus macaques at the NEPRC							
Virus family	Virus name <sup>1</sup>	Name of most closely related virus <sup>2</sup>	Percent identity nt or aa <sup>3</sup> (length, nt or aa)	Animal(s) in which viruses were detected <sup>4</sup>			
				24 weeks		64 weeks	
				SIV-	SIV+	SIV-	SIV+
Picobirnaviridae	WUHARV Dependovirus 4	Adeno-associated virus isolate rh.8R	86% nt (988/1145)	—	37, <u>40*</u>	—	—
	WUHARV Dependovirus 5	Adeno-associated virus 7	86% nt (264/307)	—	<u>26</u>	—	—
	WUHARV Parvovirus 1	Bufavirus 2	77% nt (1626/2111)	7	24*, 31, 33, 38, <u>39*</u>	—	25, 37
	WUHARV Parvovirus 2	Bufavirus 2	75-79% nt (522/698, 139/175)	—	<u>35*</u>	—	27
	WUHARV Picobirnavirus 1	Human picobirnavirus	26% aa (147/574)	—	<u>26, 36*</u>	—	—
	WUHARV Picobirnavirus 2	Otarine picobirnavirus	29% aa (119/415)	14, 11	—	—	—
	WUHARV Picobirnavirus 3	Human picobirnavirus	29% aa (102/354)	8, <u>15</u>	—	—	—
	WUHARV Picobirnavirus 4	Human picobirnavirus	36% aa (94/260)	—	<u>27</u>	—	—
	WUHARV Picobirnavirus 5	Human picobirnavirus	34% aa (63/187)	<u>22</u>	37	—	—
	WUHARV Picobirnavirus 6	Human picobirnavirus	37% aa (55/149)	—	<u>36*</u>	—	—
Picornaviridae	WUHARV Enterovirus 1	Human enterovirus 92 strain RJG7	86% nt (6228/7268)	—	<u>41, 44*</u>	—	—
	WUHARV Enterovirus 2	Simian enterovirus SV19 strain M19s	83% nt (5879/7100)	—	23, 25, 26, 27, 29, 31, 35*, 38, <u>39*</u>	—	25, 29, 30, 32, 33, 38, 41
	WUHARV Enterovirus 3	Simian enterovirus SV19 strain M19s	84% nt (5854/6961)	—	—	—	25, 26, 27, 29, 30, <u>33</u> , 37, 41
	WUHARV Picornavirus 1	Simian picornavirus 6	85% nt (284/335)	—	<u>26</u>	—	—
	WUHARV Sapelovirus 1	Simian sapelovirus 1	81% nt (6558/8087)	19	25, 26, 29, 30, 31, 32, 34*, 35*, 38, 39*, <u>42*</u> , 43*	1, 11, 17, 19, 22	23, 25, 26, 30, 31, 32, 33, 37, 38
	WUHARV Sapelovirus 2	Simian sapelovirus 1	81% nt (6510/8076)	—	25, 28*, 35*, 37, <u>41</u>	—	29, 41
	WUHARV Sapelovirus 3	Simian sapelovirus 1	79% nt (5476/6919)	—	23, <u>37</u> , 40*	—	27
	WUHARV Polyomavirus 1	Polyomavirus HPyV6 isolate 601a	76% nt (242/318)	—	—	—	29

<sup>1</sup>Viruses with 98% nt identity over the full length of aligned regions are the "same" virus.  
<sup>2</sup>Most closely related viruses were identified as the top hit using a NCBI web-based BLAST search against the NCBI nr database on Aug. 31, 2012.  
<sup>3</sup>Percent aa identity is reported for viruses for which no known virus had nt identity.  
<sup>4</sup>As determined by 454 sequencing.  
<sup>5</sup>Underlined numbers indicate animals from which virus contigs were assembled.  
\*Euthanized for progressive AIDS 24 to 64 weeks after SIV infection

Genomic Analysis of Viruses in Rhesus Monkeys at the NEPRC

**[0129]** We next analyzed the viruses present in the NEPRC cohort by assembling virus sequences from individual animals into contigs which could then be compared to the most closely related virus present in the database (see, e.g., FIGS. 5A-5D, 6A, and Table 3). Here, these viruses will be named using the convention “WUHARV-virus name-number.” Even within a single animal we found contigs from distinct but related viruses, indicating that some animals were shedding more than one virus of the same genus (FIGS. 4, 5, and Table 3).

**[0130]** We detected at least four adenoviruses (WUHARV Adenovirus 1-4) in the NEPRC cohort (FIG. 6A depicts WUHARV Adenovirus 1). We assembled portions of three calicivirus (WUHARV Caliciviruses 1-3) genomes (FIG. 5A). Importantly, WUHARV Caliciviruses 1 and 2 were most closely related to, but distinct from, the known primate calicivirus virus Tulane (Farkas et al., *J. Virol.* 82: 5408-5416 (2008); Farkas et al., *J. Gen. Virol.* 91: 734-738 (2010); Wei et al., *J. Virol.* 82: 11429-11436 (2008); Farkas et al., *J. Virol.* 84: 8617-8625 (2010)). For example, WUHARV Calicivirus 1 shared only 75% nucleotide identity over the 6,489-bp contig we assembled with Tulane calicivirus and was phylogenetically distinct from Tulane (FIG. 7A). WUHARV Calicivirus 3 was quite distantly related to either Tulane virus or to WUHARV Caliciviruses 1 and 2 (FIG. 5A). We detected parvoviruses most closely related to Bufavirus 2, a recently described parvovirus (Phan et al., *J. Virol.* [Epub ahead of print] (2012)) (FIGS. 5B and 7B). We assembled viral contigs covering most of the 7,000-8,000-bp genomes of several enteroviruses or sapeloviruses, both within the Picornaviridae (FIGS. 5C, 5D, and 7C). WUHARV Enteroviruses 1-3 share nucleotide similarity most closely with simian enterovirus SV19 with 73-84% nucleotide identity over a large portion of the genome. WUHARV Sapeloviruses 1-3 are most closely related to simian sapelovirus 1 strain 2383, sharing 79 to 81% nucleotide identity over essentially the entire genome. These data confirm that a remarkably wide variety of viruses are included within the expansion of the enteric virome associated with pathogenic SIV infection.

Next Generation Sequencing-Independent Confirmation of Virome Findings

**[0131]** We considered the possibility that relying on next generation sequencing (NGS) to document expansion of the enteric virome associated with pathogenic SIV infection might lead to false conclusions. For example, perhaps all detected viruses were present in multiple monkeys but the sequencing process is somehow biased by pathogenic SIV infection. To address this, we designed PCR assays to detect viruses for which we had large portions of the genome (FIG. 5E, Table 1) and used independent assays to detect viruses (FIG. 5E). PCR is a standard and sensitive diagnostic approach to detection of viruses in biological samples. In some cases the contigs were so divergent from each other that separate PCR assays had to be designed for different viruses in the same group. For example, one PCR assay was developed to detect WUHARV Caliciviruses 1 and 2 while a different assay was developed to detect WUHARV Calicivirus 3, which is highly divergent from all known caliciviruses. Overall PCR analysis correlated well with 454 detected viruses. PCR was positive for 454-detected viruses in 62/69 (90%) cases

(FIG. 5E), with some of the failures potentially related to the presence of viruses that were divergent from the viruses used to design PCR primers. Consistent with this, PCR detected viruses in samples when as few as 1-2 viral sequences were detected in 454-derived datasets.

**[0132]** Compared to NGS, PCR detected 5/7 adenoviruses (failing to detect divergent adenoviruses in animals #34 and #39), 14/16 caliciviruses (failing to detect divergent caliciviruses in animals #23 and #24), 10/11 parvovirus genus members (parvoviridae, failing to detect a divergent parvovirus in animal #7), 11/12 enterovirus genus members (picornaviridae, failing to detect a divergent enterovirus in animal #34) and 22/23 sapelovirus genus members (picornaviridae, failing to detect a non-divergent virus in animal #19 representing a true false negative). Importantly, PCR was negative for virus infection in a total of 151/151 cases for adenoviruses, caliciviruses, parvoviruses, enteroviruses, and sapeloviruses when next generation sequencing followed by bioinformatic analysis did not reveal a viral sequence.

**[0133]** To further confirm NGS results, we cultured viruses from fecal samples. NGS data revealed (FIGS. 4A, 6A, and Table 3) that multiple animals at NEPRC were potentially infected by novel adenoviruses. We therefore selected fecal samples from four animals for potential isolation of adenoviruses. Within the NEPRC cohort we selected feces from animals #40 (60 adenovirus sequences), #44 (138 adenovirus sequences), and #30 (2 adenovirus sequences), as well as a fourth rhesus monkey not in this cohort (57 adenovirus sequences of 5,758 unique reads) and sought to isolate viruses from them. We cultured five adenoviruses from these four animals (WUHARV Ad#1-5). These viruses were sequentially plaque purified and then isolated on cesium chloride gradients. The identity of these purified viruses as the adenoviruses detected in 454 sequencing was confirmed by PCR and sequencing (WUHARV Ad1, shown in FIG. 6A). Together both PCR and culture analyses confirmed the presence of viruses detected by NGS in fecal samples from pathogenic SIV-infected animals.

#### Example 4

##### Novel Viruses Detected by Next Generation Sequencing are Associated with AIDS Enteropathy

**[0134]** We next considered the possibility that detection of viral sequences in feces would predict intestinal disease in SIV-infected rhesus monkeys. This is a key question because our other data demonstrate only that viruses are shed in feces. To determine if viruses detected by sequencing can be clinically significant we evaluated the small and large intestine of 12 SIV-infected rhesus monkeys housed at the NEPRC (FIG. 1B, results summarized in Table 4). Of the 12 animals necropsied, six had intestinal pathology characteristic of infection with cytomegalovirus or *Balantidium* (Table 4). Importantly, analysis of the fecal virome of these 12 animals revealed that three animals (#23, 27, and 41) had high levels of adenovirus sequences prior to necropsy (Table 4). These three rhesus macaques, but not others in this necropsy cohort, exhibited adenovirus-associated enteritis by histologic examination (FIGS. 6C and 6D, (i) and (ii)). Of these three monkeys with enteropathy, all had lesions in the jejunum and ileum (ileitis) while one also had lesions in the cecum (colitis). Immunohistochemistry for adenovirus confirmed the diagnosis of adenovirus ileitis and colitis (FIGS. 6C and 6D, (iii) and (iv)). Together these data demonstrate that viruses detected in the

fecal material from SIV-infected rhesus monkeys using next generation sequencing can cause intestinal disease and epithelial damage in SIV-infected macaques.

**[0135]** To further investigate the clinical relevance of viruses detected by NGS, we used virus specific PCR assays (Table 1) to determine whether viruses detected in the fecal material of SIV-infected rhesus monkeys (FIGS. 4A and 5E) were present in serum. We detected parvovirus (FIG. 5E) in 4/10 serum samples taken at the time animals were euthanized for advanced AIDS between 24 and 64 weeks post-infection. Sequence analysis of PCR amplicons demonstrated that  $\frac{3}{4}$  viruses present in fecal material (animals #24, #28, and #39) were also present in serum. This suggests that viruses detected in the fecal material of SIV-infected rhesus monkeys can invade tissues and enter the circulation, further supporting the conclusion that SIV-associated expansion of the enteric virus may contribute to disease.

*English Edition*. Amsterdam: Elsevier Science (1998)). There was a statistically significant difference in Shannon Diversity in the NEPRC cohort sampled 64 weeks post-infection between SIV infected and uninfected control monkeys ( $p=0.0345$ ), but this was not replicated in either other cohort of monkey infected with pathogenic SIV (NEPRC cohort at 24 weeks of infection, TNPRC cohort, FIG. 8C). There were no significant differences between SIV-infected and uninfected monkeys in any cohort amongst the most-represented 20 bacterial families (FIGS. 9A-9D). There was no significant difference in bacterial family evenness across cohorts (FIG. 8C-8F). Additional analysis using principal component analysis and both supervised and unsupervised random forest analysis (Yatsuneko et al., *Nature*. 486: 222-227 (2012)) showed no association between SIV infection and the bacterial microbiome. Further we failed to find an association between SIV infection and either the genus- or species-level

TABLE 3

Summary of adenovirus detection and pathology in SIV-infected rhesus monkeys							
Animal number	Adenovirus reads <sup>a</sup>	WUHARV Adenovirus <sup>a</sup>	PCR screen <sup>a,b</sup>	Adenovirus Enteritis <sup>c</sup>	SI Adenovirus IHC <sup>c</sup>	LI Adenovirus IHC <sup>c</sup>	Other GI Pathologies <sup>c</sup>
23	889	1, others <sup>d</sup>	Pos	Yes	Pos	Neg	<i>Cytomegalovirus</i> enteritis
25	0	n/a	Neg	No	Neg	Neg	No
26	0	n/a	Neg	No	Neg	Neg	No
27	653	5, others <sup>d</sup>	Pos	Yes	Pos	Neg	<i>Balantidium</i> sp typhilitis
29	14	others <sup>d</sup>	Neg	No	Neg	Neg	No
30	1	others <sup>d</sup>	Neg	No	Neg	Neg	No
31	0	n/a	Neg	No	Neg	Neg	<i>Mycobacterium avium</i> enteritis; Ⓣ
32	52	others <sup>d</sup>	Neg	No	Neg	Neg	No
33	4	others <sup>d</sup>	Neg	No	Neg	Neg	No
37	0	n/a	Neg	No	Neg	Neg	No
38	0	n/a	Neg	No	Neg	Neg	<i>Balantidium</i> sp colitis
41	640	others <sup>d</sup>	Pos	Yes	Pos	Pos	<i>Balantidium</i> sp

<sup>a</sup>Number of adenovirus sequences detected at 64 weeks.

<sup>b</sup>Results from PCR for indicated adenovirus (primers, Supplemental Table 1).

<sup>c</sup>Results obtained at necropsy.

<sup>d</sup>Novel adenoviruses highly diverged from Adenovirus 1-5 as well known adenoviruses.

Ⓣ indicates text missing or illegible when filed

### Example 5

#### SIV Infection and the Bacterial Microbiome

**[0136]** We next assessed the effects of SIV infection on the taxonomy of the bacterial microbiome. Our metagenomic data was comparable to published 16S rDNA-derived class-level data from SIV-infected and control macaques at TNPRC (McKenna et al., *PLoS*. 4: e20 (2008)), indicating that these distinct methods yield overall similar results (FIGS. 8A and 8B). Rarefaction analysis revealed that all but a few samples with very high numbers of viral sequences were robust for analysis of bacterial diversity at the family level (FIGS. 8C-8F). Species accumulation curves indicated that all cohorts except the NEPRC African green monkey cohort were robust for this analysis; further analysis excluded this cohort (FIG. 8G). We detected no consistent SIV-associated differences in bacterial family richness, evenness, or diversity (Legendre and Legendre. *Numerical Ecology*, *Second*

taxonomic structure of the bacterial microbiome. Thus, in contrast to our analysis of the virome, we detected no consistent SIV-infection associated differences in the family-level taxonomy of the bacterial microbiome.

#### OTHER EMBODIMENTS

**[0137]** While the invention has been described in connection with specific embodiments thereof, it will be understood that it is capable of further modifications and this application is intended to cover any variations, uses, or adaptations of the invention following, in general, the principles of the invention and including such departures from the present disclosure that come within known or customary practice within the art to which the invention pertains and may be applied to the essential features hereinbefore set forth.

**[0138]** All publications and patent applications mentioned in this specification are herein incorporated by reference to the same extent as if each independent publication or patent application was specifically and individually indicated as being incorporated by reference in their entirety.

## SEQUENCE LISTING

The patent application contains a lengthy "Sequence Listing" section. A copy of the "Sequence Listing" is available in electronic form from the USPTO web site (<http://seqdata.uspto.gov/?pageRequest=docDetail&DocID=US20150284814A1>). An electronic copy of the "Sequence Listing" will also be available from the USPTO upon request and payment of the fee set forth in 37 CFR 1.19(b)(3).

What is claimed is:

**1-32.** (canceled)

**33.** A method of detecting acquired immune deficiency syndrome (AIDS) and/or AIDS progression in a subject, said method comprising:

- a) synthesizing cDNA from RNA by a biological sample obtained from as subject; and
- b) detecting if present in the sample one or more target nucleotide sequences that specifically hybridize under stringent conditions to one or more polynucleotides having at least 70% sequence identity with any one of SEQ ID NO:1-107 by a PCR assay selected from the group consisting of a real time PCR assay and a nested PCR assay; and

determining the magnitude of difference between the level of said one or more target nucleotide sequences in said subject relative to the level of said one or more target nucleotide sequences from a control subject whereby an increase in the level of said one or more target nucleotide sequences in said subject, relative to the level of the one or more target nucleotide sequences from a control subject, indicates AIDS and/or AIDS progression in said subject.

**34.** A method of diagnosing, or providing a prognostic indicator of, immunodeficiency and/or enteropathy in a subject, said method comprising:

providing at least one biological sample from the subject; subjecting RNA if present from the at least one biological sample to synthesis of cDNA comprising one or more target nucleotide sequences that specifically hybridize under stringent conditions to one or more of polynucleotides having a sequence that is at least 70% identical to the sequence of a polynucleotide selected from the group consisting of SEQ ID NO:1-107;

detecting said one or more target nucleotide sequences by a PCR assay selected from the group consisting of a real time PCR assay and a nested PCR assay; and

determining the magnitude of difference between the level of said one or more target nucleotide sequences in said subject relative to the level of one or more target nucleotide sequences from a control subject wherein the detection of an increase in the level of said one or more target nucleotide sequences in said subject, relative to the level of one or more target nucleotide sequences from a control subject, indicates the presence of, or the propensity to develop, immunodeficiency and/or enteropathy in said subject.

**35.** A method of identifying a therapeutic agent for treating immunodeficiency and/or enteropathy in a subject, said method comprising:

providing a first sample from the subject prior to administration of a candidate therapeutic agent; administering a therapeutically effective amount of the candidate therapeutic agent to the subject; providing a second sample from the subject after the administering the candidate therapeutic agent; detecting in the first sample by a first PCR assay one or more target polynucleotides that specifically hybridize under stringent conditions to one or more polynucleotides having a sequence that is at least 70% identical to the sequence of a polynucleotide selected from the group consisting of SEQ ID NO: 1-107; detecting in the second sample by a second PCR assay the one or more target polynucleotides, wherein the detection of a decrease in the level of said one or more target nucleotide sequences in said second sample, relative to the level of the one or more target nucleotide sequences from said first sample, identifies said candidate agent as said therapeutic agent.

**36.** (canceled)

**37.** The method of claim **35**, wherein said immunodeficiency or enteropathy is associated with a lentivirus.

**38.** The method of claim **37**, wherein said immunodeficiency and enteropathy are associated with a lentivirus.

**39.** The method of claim **37**, wherein said lentivirus is human immunodeficiency virus (HIV) or simian immunodeficiency virus (SIV).

**40.** The method of claim **35**, wherein the detecting in the first sample one or more target nucleotide sequences comprises synthesizing cDNA from RNA comprised by said sample.

**41.** (canceled)

**42.** The method of claim **40**, wherein said first PCR assay is selected from the group consisting of a real time PCR (RT-PCR) assay and a nested PCR assay.

**43.** The method of claim **33**, wherein said sample is a tissue, organ, liquid, or feces sample.

**44.** The method of claim **43**, wherein said subject is a mammal.

**45.** The method of claim **44**, wherein said mammal is a primate.

**46.** The method of claim **45**, wherein said primate is a human.

**47.** The method of claim **38**, wherein said lentivirus is human immunodeficiency virus (HIV).

**48.** The method of claim **38**, wherein said lentivirus is simian immunodeficiency virus (SIV).

**49.** The method of claim **34**, wherein said immunodeficiency or enteropathy is associated with a lentivirus.

**50.** The method of claim **49**, wherein said lentivirus is human immunodeficiency virus (HIV).

**51.** The method of claim **49**, wherein said lentivirus is simian immunodeficiency virus (SIV).

**52.** The method of claim **33**, wherein said sample is a feces sample.

**53.** The method of claim **34**, wherein said sample is a feces sample.

**54.** The method of claim **35**, wherein said first sample is a first feces sample and said second sample is a second feces sample.

\* \* \* \* \*