

Gene Therapy Vectors Based on Adeno-Associated Virus Type 1

WEIDONG XIAO, NARENDRA CHIRMULE, SCOTT C. BERTA, BETH McCULLOUGH,
GUANGPING GAO, AND JAMES M. WILSON*

*Institute for Human Gene Therapy and Departments of Molecular and Cellular Engineering
and of Medicine, University of Pennsylvania, and The Wistar Institute, Philadelphia, Pennsylvania 19104*

Received 15 December 1998/Accepted 29 January 1999

The complete sequence of adeno-associated virus type 1 (AAV-1) was defined. Its genome of 4,718 nucleotides demonstrates high homology with those of other AAV serotypes, including AAV-6, which appears to have arisen from homologous recombination between AAV-1 and AAV-2. Analysis of sera from nonhuman and human primates for neutralizing antibodies (NAB) against AAV-1 and AAV-2 revealed the following. (i) NAB to AAV-1 are more common than NAB to AAV-2 in nonhuman primates, while the reverse is true in humans; and (ii) sera from 36% of nonhuman primates neutralized AAV-1 but not AAV-2, while sera from 8% of humans neutralized AAV-2 but not AAV-1. An infectious clone of AAV-1 was isolated from a replicated monomer form, and vectors were created with AAV-2 inverted terminal repeats and AAV-1 Rep and Cap functions. Both AAV-1- and AAV-2-based vectors transduced murine liver and muscle *in vivo*; AAV-1 was more efficient for muscle, while AAV-2 transduced liver more efficiently. Strong NAB responses were detected for each vector administered to murine skeletal muscle; these responses prevented readministration of the same serotype but did not substantially cross-neutralize the other serotype. Similar results were observed in the context of liver-directed gene transfer, except for a significant, but incomplete, neutralization of AAV-1 from a previous treatment with AAV-2. Vectors based on AAV-1 may be preferred in some applications of human gene therapy.

Adeno-associated viruses (AAV) are small, nonenveloped, single-stranded DNA viruses which require helper virus to facilitate efficient replication (3). The 4.7-kb genome of AAV is characterized by two inverted terminal repeats (ITRs) and two sets of open reading frames, which encode the Rep and Cap proteins. The Rep open reading frames encode four proteins with molecular masses of 78, 68, 52, and 40 kDa. These proteins function mainly in regulating AAV replication and integration. The Cap open reading frames encode three structural proteins with molecular masses of 85 kDa (VP1), 72 kDa (VP2), and 61 kDa (VP3) (3). The two ITRs are the only *cis* elements essential for all steps in the AAV life cycle.

AAV have been found in many animal species, including nonhuman primates, canines, fowl, and humans (18). A total of six serotypes of AAV, including AAV type 1 (AAV-1), have been isolated from primates, and two have been isolated from nonhuman primates; AAV-2, AAV-3, and AAV-5 are from humans, and AAV-6 is from a human adenovirus preparation. AAV-2 is the most characterized primate serotype, since its infectious clone was the first one made (24). The full sequences for AAV-3A, AAV-3B, AAV-4, and AAV-6 recently were determined (4, 17, 22). Generally, all primate AAV show more than 80% homology in nucleotide sequence.

A number of unique properties make AAV a very promising vector for human gene therapy (19). AAV are not associated with any known human diseases and are generally not considered pathogenic. Wild-type AAV are capable of integrating into the host chromosome in a site-specific manner (14, 26). Recombinant AAV (rAAV) vectors can integrate into tissue culture cells at chromosome 19 if the Rep proteins are supplied *in trans* (1, 29). The transduced genomes of AAV have been shown to confer long-term gene expression in a number of tissues, including muscle, liver, brain, and retina (8, 13, 27, 28,

30, 31). The development of new methods for producing high-titer rAAV has largely removed the hurdles which prevented AAV vectors from being tested in large-animal models of human diseases and in human clinical trials (5, 6, 11, 32).

Among AAV-1 to AAV-6, only AAV-1 and AAV-4 are considered to be simian viruses, since they were isolated from nonhuman primates and monospecific antibodies to the viruses have not been detected in human serum (20). They may have advantages for use in human gene therapy to replace or augment the use of AAV-2 vectors. For example, AAV-1 vectors could be used in patients who develop anti-AAV-2 neutralizing antibodies (NAB) due to a naturally acquired infection or previous treatment with AAV-2 vectors. To study the possibility of using AAV-1 as a gene therapy vector, we constructed an AAV-1 infectious clone and determined its full sequence. Vectors derived from this infected clone were evaluated in murine models of liver and muscle gene transfer.

MATERIALS AND METHODS

Murine studies. C57BL/6 mice (6- to 8-week-old males) were obtained from Jackson Laboratory. AAV vectors were administered by either intramuscular or intrasplenic injection as described before (8, 30).

Nonhuman primates. Wild-caught juvenile rhesus monkeys were purchased from Covance (Alice, Tex.) and LABS of Virginia (Yemassee, S.C.) and kept in full quarantine. The monkeys weighed approximately 3 to 4 kg. The nonhuman primates used in the Institute for Human Gene Therapy research program are purposefully bred in the United States from specific-pathogen-free closed colonies. All vendors are U.S. Department of Agriculture class A dealers. The rhesus macaques are therefore not infected with important simian pathogens, including the tuberculosis agent, major simian lentiviruses (simian immunodeficiency virus and simian retroviruses), and cercarial dermatitis. The animals are also free of internal and external parasites. The excellent health status of these premium animals minimizes the potential for extraneous variables. The use of these monkeys in protocols was approved by the Infection Control Committee of the Hospital of the University of Pennsylvania and the Environmental Health and Safety Office, the Institutional Biosafety Committee, and the Institutional Animal Care and Use Committee of the University of Pennsylvania. For this study, serum was obtained from monkeys prior to initiation of any protocol.

Human subjects. Normal volunteers (*n*, 77) were analyzed for immune reactivity to AAV. Individuals were members of the University of Pennsylvania community. The study was approved by the Institutional Review Board of the

* Corresponding author. Mailing address: 204 Wistar Institute, 3601 Spruce St., Philadelphia, PA 19104-4268. Phone: (215) 898-3000. Fax: (215) 898-6588. E-mail: wilsonjm@mail.med.upenn.edu.

University of Pennsylvania. The median age was 27 years (range, 18 to 54 years), and the age distributions in the two groups were similar.

Cell culture and virus. AAV-free 293 cells and 84-31 cells were obtained from the Human Applications Laboratory of the University of Pennsylvania. These cells were cultured in Dulbecco's modified Eagle medium with 10% fetal bovine serum (Hyclone), penicillin (100 U/ml), and streptomycin (100 µg/ml) at 37°C in a moisturized environment supplied with 5% CO₂. The 84-31 cell line constitutively expresses adenovirus E4 proteins and has been described previously (7). AAV-1 (ATCC VR-645) seed stock was purchased from the American Type Culture Collection (Rockville, Md.). AAV were propagated in 293 cells with wild-type adenovirus type 5 as a helper virus.

rAAV generation. rAAV were generated by transfection with an adenovirus-free method described elsewhere (30a). Briefly, the *cis* plasmid (with AAV ITRs), the *trans* plasmid (with the AAV *Rep* gene and *Cap* gene), and a helper plasmid (pΔ13, with an essential region from the adenovirus genome) were cotransfected into 293 cells at a ratio of 1:1:2 by calcium phosphate precipitation. The AAV vectors used in this study express murine erythropoietin or human α1-antitrypsin from a chicken β-actin promoter enhanced by sequences from cytomegalovirus. pΔ13 has an 8-kb deletion in the adenovirus E2B region and most of the late genes. The cells were harvested 96 h posttransfection and subjected to two rounds of CsCl gradient purification. For the generation of rAAV based on AAV-2, p5E18 was used as the *trans* plasmid, since it greatly improved the rAAV yield. To make rAAV with AAV-1 virions, pAV1H or p5E18(2/1) was used as the *trans* plasmid to provide *Rep* and *Cap* functions.

DNA techniques. Hirt DNA extraction was performed as described before with minor modifications (25).

To construct AAV-1 infectious clones, the Hirt DNA from AAV-1-infected 293 cells was repaired with the Klenow enzyme (New England BioLabs) to make sure that the ends were blunt. The treated AAV-1 Hirt DNA was then digested with *Bam*HI and cloned into three vectors. The internal *Bam*HI fragment was cloned into pBluescript II-SK(+) cut with *Bam*HI to produce pAV1-BM. The left and right fragments were cloned into pBluescript II-SK(+) cut with *Bam*HI and *Eco*RV to obtain pAV1-BL and pAV1-BR, respectively. The AAV sequences in these three plasmids were subsequently assembled into the same vector to produce AAV-1 infectious clone pAAV-1.

For sequencing, an ABI 373 automatic sequencer was used to determine the sequences for all plasmids and PCR fragments used in this study by fluorescence sequencing (FS) dye chemistry. They were confirmed by sequencing both plus and minus strands. For sequencing the AAV terminal repeats, the fragment containing AAV ITRs was amplified from pAAV-1, pAV1-BR, or pAV1-BL with M13F (GTAAAACGACGGCCAGT) and CAP1a (GAA CCG TGC TGG TAA GGT TAT T) or M13F and V1Rep1s (CCA TGC CGG GCT TCT ACG AGA TCG TTA TCA GGG TG). The PCR was performed with a GC-advance kit (Clontech Laboratories, Inc.). The 50-µl reaction mixture contained 1× reaction buffer (40 mM Tricine-KOH [pH 9.02] at 25°C, 15 mM KOAc, 3.5 mM Mg(OAc)₂, 5% dimethyl sulfoxide, 3.75 µg of bovine serum albumin per ml); 1× GC melt; 10 mM dATP, dCTP, 7-deazaGTP-dGTP (3:1), and dTTP; 10 ng of template; 500 ng of each primer; and 2.5 U of polymerase. The purified PCR product was sequenced as described above.

A plasmid expressing AAV-2 *Rep* and *Cap*, p5E18(2/2), was constructed to maximize vector production. This was accomplished by relocating the p5 promoter to a position 3' to the *Cap* gene, thereby minimizing expression of *Rep*78 and *Rep*68. This strategy was initially described by Li et al. (15). p5E18(2/2) was constructed in the following way. The previously described pMTV-trans vector (30a) (i.e., the mouse mammary tumor virus promoter substituted for the p5 promoter in an AAV-2-based vector) was digested with *Sma*I and *Clal*, filled in with the Klenow enzyme, and then recircularized with DNA ligase. The resulting construct was digested with *Xba*I, filled in, and ligated to the blunt-ended *Bam*HI-*Xba*I fragment from pCR-p5, constructed in the following way. The p5 promoter of AAV was amplified by PCR with the following oligonucleotides: TGT AGT TAA TTA ACC CGC CAT GCT ACT TAT C and GGC GGC TGC GCG TTC AAA CCT CCC GCT TCA AAA TG. The amplified fragment was subsequently cloned into pCR2.1 (Invitrogen) to yield pCR-p5. The helper plasmid pAV1H was constructed by cloning the *Bfa*I fragment of pAAV-1 into pBluescript II-SK(+) at the *Bco*RV and *Sma*I sites. The 3.0-kb *Xba*I-*Kpn*I fragment from p5E18(2/2), the 2.3-kb *Xba*I-*Kpn*I fragment from pAV1H, and the 1.7-kb *Kpn*I fragment from p5E18(2/2) were incorporated into a separate plasmid, p5E18(2/1), which contains AAV-2 *Rep*, AAV-1 *Cap*, and the AAV-2 p5 promoter located 3' to the *Cap* gene.

For Southern blot analysis, Hirt DNA was digested with *Dpn*I to remove bacteriophage-borne plasmid and probed with the internal *Bam*HI fragment of AAV-1. The membrane was then washed under high-stringency conditions: twice for 30 min each time with 2× SSC (1× SSC is 0.15 M NaCl plus 0.015 M sodium citrate)-0.1% sodium dodecyl sulfate (SDS) at 65°C and twice for 30 min each time with 0.1× SSC-0.1% SDS at 65°C. The membrane was then analyzed by both phosphorimager analysis and X-ray autoradiography.

Western blot analyses. Serum samples were analyzed for reactivity to various AAV-2 capsid proteins by Western blotting as described previously (8). Purified AAV-2 vectors were used as antigens for the capsid proteins, and a monoclonal antibody that recognizes shared epitopes on VP1, VP2, and VP3 (clone B1; American Research Products) was used as a positive control. AAV-2 antigens were electrophoresed by SDS-10% polyacrylamide gel electrophoresis and trans-

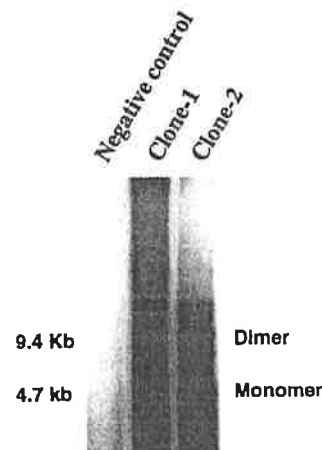


FIG. 1. DNA hybridization analysis of the rescue and replication of the AAV-1 infectious clone. Two independent clones of pAAV-1 (Clone-1 and Clone-2) were transfected into 293 cells, which were then superinfected with adenovirus type 5. Hirt DNA was extracted 48 h postinfection and digested with *Dpn*I before electrophoresis in an 0.8% agarose gel. An internal *Bam*HI fragment of pAAV-1 was used as a probe. The autoradiograph shows the locations of replicated dimers and monomers and their molecular sizes. The analysis included cells not infected (negative control) or infected with clone-1 or clone-2.

ferred to a nitrocellulose membrane (Hybond-ECL; Amersham). The reactivity of the serum was measured by incubating membrane containing AAV antigen with test serum samples, followed by peroxidase anti-human immunoglobulin G antibody. The reaction was detected with an ECL kit (Amersham).

Anti-AAV NAB. NAB titers were analyzed by assessing the ability of serum antibody to inhibit the transduction of reporter virus (AAV1-GFP or AAV2-GFP) into 84-31 cells, which are subclones of 293 cells that stably express adenovirus E4 proteins, which render them permissive for AAV transduction. Various dilutions of antibodies preincubated with reporter virus for 1 h at 37°C were added to 90% confluent cell cultures. Cells were incubated for 48 h, and the expression of green fluorescent protein (GFP) was measured by FluorImaging (Molecular Dynamics). NAB titers were calculated as the highest dilution at which 50% of the cells stained green.

Human α1-antitrypsin assay. The concentration of human α1-antitrypsin in mouse serum was measured by an enzyme-linked immunosorbent assay (ELISA). The coating antibody was rabbit anti-human α1-antitrypsin (Sigma). We used goat anti-human α1-antitrypsin (Sigma) as the primary detection antibody (30). The sensitivity of the assay is 0.3 to 30 ng/ml.

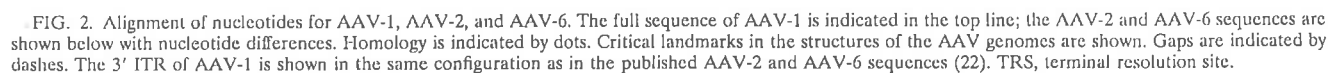
Nucleotide sequence accession number. The AAV-1 sequence is available through GenBank under accession no. AF063497.

RESULTS

Generation of an infectious clone of AAV-1. The replicated form of AAV-1 was extracted from 293 cells infected with AAV-1 and wild-type adenovirus type 5. Analysis of Hirt DNA revealed three bands with apparent molecular sizes equivalent to a double-stranded dimer (9.4 kb), a double-stranded monomer (4.7 kb), and single-stranded DNA (1.7 kb). The monomer band was excised from the gel and digested with *Bam*HI, revealing a pattern of three fragments (1.1, 0.8, and 2.8 kb), in accordance with the original description of AAV-1 by Bantel-Schaal and zur Hausen (2). The three fragments were subcloned into one plasmid-based construct to obtain pAAV-1.

This clone was tested for its abilities to rescue AAV-1 genomes from the plasmid backbone and to replicate and package infectious virus. Plasmid pAAV-1 was transfected into 293 cells infected with adenovirus type 5 at a multiplicity of infection of 10, and the virus supernatant was used to reinfect 293 cells. DNA hybridization analysis of *Dpn*I-digested Hirt DNA from infected 293 cells demonstrated the rescue and replication of AAV-1 genomes (Fig. 1).

Genomic structure of AAV-1. The entire AAV-1 genome was sequenced from pAAV-1 (Fig. 2). PCR products derived



The AAV-1 genome shows more than 80% identity with other known AAV and contains the characteristic structural

AAV-1	CCGACAGGTACCAAAACAAATGTTCTCGTCACGCGGGCATGCTTCAGATGCTGTTCCCTGCAAGACATGCGAGAGAATGAATCAGAATTTCAACATTTGCTTCACGCACGGGACGAGAG	2036
AAV-2	..A.....T.....AA..T.....GACA.....CA..T..C.....T.....ACA..A..	2019
AAV-6	A.....C.....	2021
AAV-1	ACTGTTCAGAGTCTTCCCGCGTGTCTCAGAACTCAACCGGTC---GTCAGAAAGAGGACGTATCGGAACTCTGTGCCATTATCATCTGCTGGGGCGGGCTCCCGAGATTGCTTGCT	2153
AAV-2	T.....T.....C..TTCT...GTC...A..A..G.....A.....G..CTA.....A..CA.....AAA..TG..A.....C.....A	2133
AAV-6	A..T.....	2138
AAV-1	CGGCTCGCATCTGGTCAACGTGGACCTGGATGACTGTGTTTCTGAGCAATAATGACTTAAACCGGATATGGCTGCCGATGTTATCTTCAGATTGCTCGAGGACAACTCTCTGAG	2273
AAV-2	..T.....T.....TT.....CA..C..T...A.....T.....T.....CT.....A	2253
AAV-6	T.....AC.....G	2258
AAV-1	GGCATTTCGAGTGTGGGACTTGAACCTGGAGCCCGCAAGCCCAAGCCAAAGCAGCAAGAGCGGCGGGTCTGGTGCTTCTGGGTACAAAGTACCTCGGACCCCTTCAAC	2393
AAV-2	..A..AA..AC.....A..GC..C.....CC..A..ACCA..A..GC..GCAAG...GGC..TA.....A..A.....T.....G.....	2373
AAV-6	A.....G..C.....G.....C.....	2378
AAV-1	GGACTCGAAGGGGGAGCCCGTCAACGCGGGGAGCAGCGGCCCTCGAGCAAGCAAGGCGCTACGACGAGCAGCTCAAGCGGGTGCAATCCGTACCTGCGGTATAACACCGCGAC	2513
AAV-2	A.....G.....A.....A.....C.....A.....G.....G..CAGC..A.....C.....CAA...C.....	2493
AAV-6	T.....A..AGCG..T.....T.....CGC...T.....	2498
AAV-1	GCCGAGTTTCAGGAGCGTCTGCAAGAGATACGTCTTTTGGGGCAACCTCGGGCGAGCAGTCTTCAGGCCAAGAGCGGGTCTCGAACCTCTCGGTCTGGTTGAGGAAGCGCTAAG	2633
AAV-2	..G.....C..TA.....A.....G..A.....T.....G..C.....CCT..T.....	2613
AAV-6	..C.....T..GC.....G.....A.....T.....T.....	2628
AAV-1	ACGGCTCCTGGAAAGAAACCTCCGGTAGAGCAGTCCGCCAAGAGCCAGACTCTCTCTCGGGCATCGGCAAGACAGGCCAGCAGCCCGCTAAGAGAGACTCAATTTTGGTCAGACTGGC	2753
AAV-2	G.....A..GA..G.....C..T..TGTG.....A..C.....A.....G..G.....T..A..G.....A.....T..G.....A	2733
AAV-6	T.....G..AC..T.....G..G..ACAA.....T.....	2738
AAV-1	GACTCAGAGTCAGTCCCGATCCCAACCTCTCGGAGAACCTCCAGCAACCCCGCTGCTGTGGGACCTACTACAATGGCTTCAGGCGGTGGCGACCAATGGCAGACATAACGAAGGC	2873
AAV-2	..G..C.....A..T..C..C..G.....C..G..A.....G.....T..G..C.....A..A.....G.....A.....A.....G.....	2853
AAV-6	..T..G.....C..C..C..A..A.....G..A..T.....A.....G.....	2858
AAV-1	GCCGACGGAGTGGGTAATGCTCAGGAAATGGCATTGCGATTCCATGGCTGGGCGACAGAGTCATCACCAACGACACCCGACCTGGGCTTGCCACCTACAATAACACCTCTAC	2993
AAV-2	T.....C.....A.....C.....C.....	2973
AAV-6	A.....A.....T..C.....	2978
AAV-1	AAGCAAAATCTCCAGTGCTTCAACGGGGGCGAGCAACGACCACTACTCTGGGCTACAGCACCCCTGGGGGTATTTTGATTTCACAGATTCCACTGCCACTTTTACCACGTGACTGG	3113
AAV-2	..A.....T.....CCAA.....A..TCG.....T.....T.....T.....C.....	3090
AAV-6	T.....C.....	3098
AAV-1	CAGCGACTCATCAACAAATTGGGGATTCCGGCCCAAGAGACTCAACTTCAAACTCTTCAACATCCAAGTCAAGGAGGTCCAGCAAGATGATGGCGTCACAAACATCGCTAATAACCTT	3233
AAV-2	..AA.....C.....A.....G.....T.....T.....A..CG..G..CA.....C..TACG..G..G..T..C.....	3210
AAV-6	G.....	3218
AAV-1	ACCAGCACGGTTCAAGTCTTCTCGGACTCGGAGTACCAGCTTCCGTACGCTCTGGCTCTGGCCACCAGGGCTGCCTCCCTCCGTTCCCGGCGGACGTGTTTCATGATTCCGCAATACGGC	3353
AAV-2	G..G..TA..T.....A.....A.....A.....C.....G.....G..A.....A.....T..A	3330
AAV-6	T..G.....	3338
AAV-1	TACCTGACGCTCAACAAATGGCAGCCAAAGCCGTGGGACGTTTCATCTTTTACTGCTGGAATATTTCCCTTCTCAGATGCTGAGAACGGGCAACAACTTTACCTTCAGCTACACCTTTGAG	3473
AAV-2	C..C..G.....C..G..T..G..A..A.....C..T..A.....C..T..C.....C..T..C.....A..A..CA..G.....TCTG..G..T.....	3450
AAV-6	A.....G..A.....G.....A..G.....T.....C.....	3458
AAV-1	GAAAGTGCCTTTCACAGCAGCTACGCGCACAGCCAGAGCCTGGACCGGCTGATGAATCCTCTCATCGACCAATACCTGTATTACCTGAACAGAACTCAAAATCAGTCCGGAAGTGCCCAA	3593
AAV-2	..C..T.....T.....T.....T.....T..C.....TC..GG..G..T..G..GC..C.....T.....AA..C..C..CAAGT.....CCA..ACG	3570
AAV-6	..C.....G.....	3578
AAV-1	AACAGGACTTGTCTTTAGCCGTGGGTCTCCAGCTGGCATGTCTGTTACGCCCAAACTGGCTACCTGGACCTGTATTTCGGCAGCAGCGCGTTTCTAAAAAAGAGACAGACAAAC	3713
AAV-2	..C..GTCAAGGC..T..A.....TCT..AG..CCGGAG..GAG..A.....TCGG..AC...T..T..GG.....T.....C..C.....A..A..CA..G.....TCTG..G..T.....	3690
AAV-6	G.....C.....	3698
AAV-1	AACAGCAATTTTACCTGGACTGGTGTCTTCAAAATATAACCTCAATGGCGTGAATCCATCATCAACCTTGGCACTGCTATGGGCTCACAAAGACGAGCAAGAGCAAGTTCTTTCCCATG	3833
AAV-2	TG..A..ACT..G.....A..C..G..CC.....CA..A..C..TC..GG..G..T..G..GC..C..C.....AAGC.....G.....T.....A.....T.....TCA..	3810
AAV-6	C.....T.....T.....A.....	3818

FIG. 2—Continued.

features. The ITRs of AAV-1 are predicted to form T-shaped hairpin structures (Fig. 3). The right and left ITRs of AAV-1 are identical and virtually the same as the right ITR of AAV-6, except for 1 nucleotide in the A and A' sequences and the

last nucleotide in the D sequence. The AAV-2 Rep binding motif (GCTCGCTCGCTCGCTG) found in the AAV-2 pre-integration region in human chromosome 19 is well conserved in AAV-1.

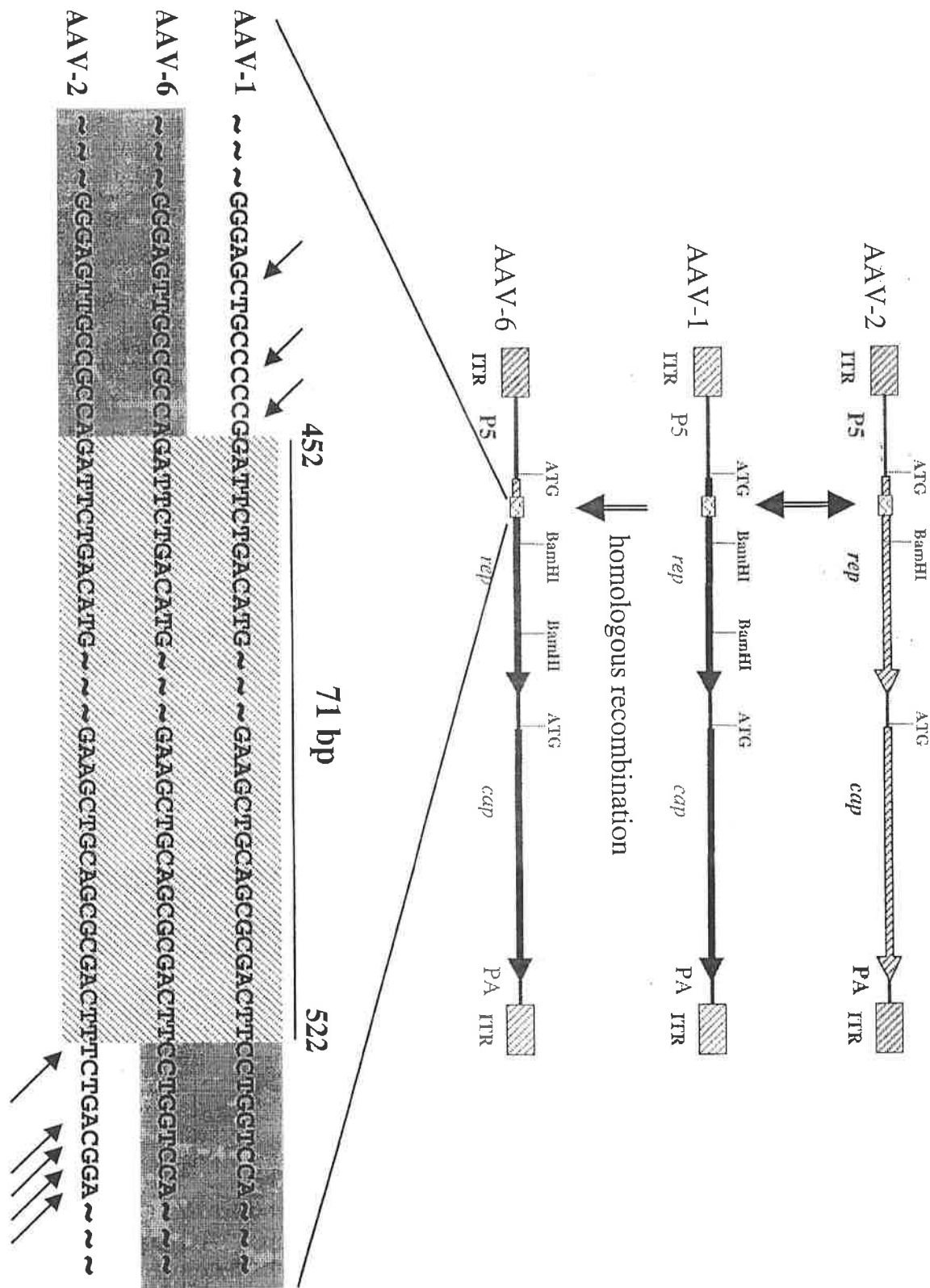


FIG. 4. Hypothesis for the creation of AAV-6 from homologous recombination between AAV-1 and AAV-2. The first line represents the AAV-2 genome, with Rep and Cap sequences demarcated by hatched arrows. The second line represents the AAV-1 genome, with the Rep and Cap sequences depicted as solid arrows. The hypothesis suggests homologous recombination between AAV-2 and AAV-1 at the region of the box containing wavy lines: this region contains a highly homologous sequence of the Rep gene shared by AAV-1 and AAV-2. A more detailed illustration of the common region of these three viruses is shown at the bottom. Arrows indicate nucleotides that differ.

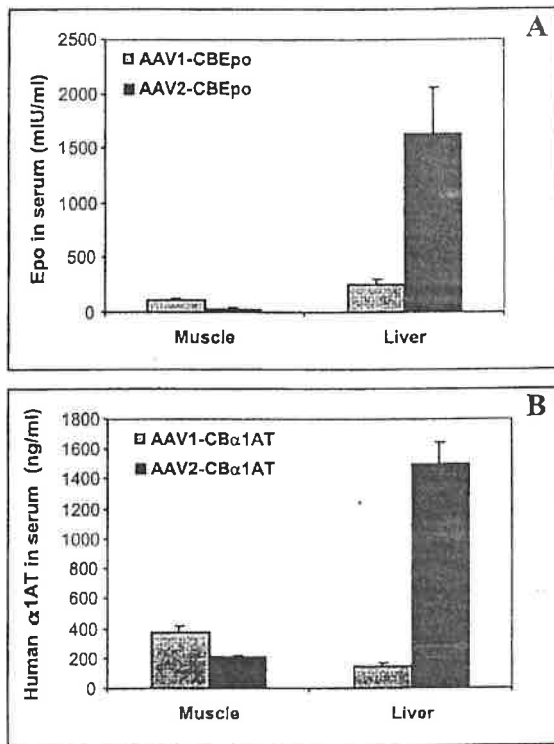


FIG. 5. In vivo activities of AAV-1 and AAV-2 vectors. Recombinant stocks of virus were generated by transfection and purified through cesium chloride gradients. (A) AAV-1 and AAV-2 vectors expressing murine erythropoietin (Epo) from a cytomegalovirus-enhanced β -actin promoter (CB). Equivalent stocks of these two vectors were injected into muscle (5×10^{10} genomes) or liver via the portal circulation (1×10^{11} genomes), and the animals (four groups) were analyzed on day 30 for the presence of erythropoietin in blood. Data for the groups were combined. Error bars show standard errors. (B) An identical experiment was performed with AAV-1 and AAV-2 vectors expressing human α 1-antitrypsin (α 1AT) from the same promoter. Serum harvested from four groups of animals 30 days after gene transfer was analyzed for the presence of α 1-antitrypsin by an ELISA. Data are reported as in panel A.

were purified on CsCl gradients, yielding low titers of the recombinant vector (i.e., 5×10^{10} genomes/50 15-cm² plates).

The AAV helper vector was revised in two ways to improve yield. First, the AAV-1 Rep gene in pAV1H was substituted for the corresponding gene in AAV-2. Second, the p5 promoter was relocated behind the Rep and Cap open reading frames to reduce the expression of the p78 and p68 forms; this modification has been used to enhance the titers of AAV-2 vectors (15,300). The resulting plasmid (p5E182/1) produced 10- to 20-fold higher quantities of the vector than pAV1H (i.e., 10^{12} genomes/50 15-cm² plates).

The performance of AAV-1 vectors was evaluated with immunodeficient mice (i.e., RAG-1 KO) injected intramuscularly or in the portal circulation to target the liver (Fig. 5). Direct comparisons were made to equivalent quantities of AAV-2 vectors containing the same vector genomes encoding secreted proteins easily measured in blood (i.e., murine erythropoietin and human α 1-antitrypsin). Early experiments indicated similar in vivo performances of AAV-1 vectors produced with pAV1H and p5E18(2/1) (data not shown). All subsequent studies used AAV-1 vectors derived from p5E18(2/1) because of the increased yield.

AAV-2 vectors consistently produced 10- to 50-fold more serum erythropoietin or α 1-antitrypsin when injected into liver compared to muscle. This result was very different from that for AAV-1 vectors, with which muscle expression was equivalent to or greater than liver expression. In fact, AAV-1 outperformed AAV-2 in muscle when equivalent titers based on genomes were administered.

NAB to AAV-1 and AAV-2 in nonhuman and human primates. Simple and quantitative assays for NAB to AAV-1 and AAV-2 were developed with recombinant vectors. A total of 33 rhesus monkeys and 77 normal human subjects were screened. Figure 6 summarizes the results. A reciprocal dilution of NAB equal to 20 is the baseline (i.e., no detectable NAB).

Analysis of NAB in rhesus monkeys showed that 61% of animals tested positive for AAV-1; a minority (24%) had NAB to AAV-2. Over one-third of animals had antibodies to AAV-

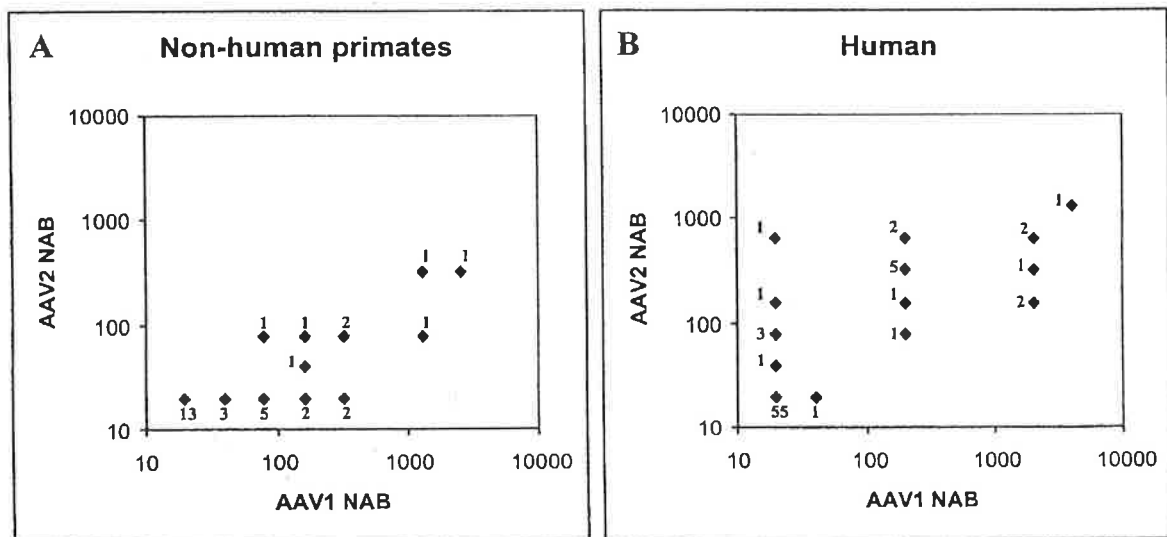


FIG. 6. NAB to AAV-1 and AAV-2 in nonhuman and human primates. Sera were analyzed for neutralizing activity against rAAV-1 and rAAV-2 as described in Materials and Methods. The reciprocal dilution of the NAB was plotted for AAV-2 versus AAV-1. Sera were titrated in sequential twofold dilutions. A titer of 1:20 represents the background (i.e., no detectable NAB). The number next to each datum point represents the total number of animals or humans with a particular NAB profile. In panel A, 33 adolescent rhesus monkeys were analyzed. In panel B, 77 normal human volunteers were evaluated.

1 but not AAV-2 (i.e., were monospecific for AAV-1), whereas no animals were positive for AAV-2 without reacting to AAV-1. These data support the hypothesis that AAV-1 is endemic in rhesus monkeys. The presence of true AAV-2 infections in this group of nonhuman primates is less clear, since we cannot rule out cross-neutralizing activity of an AAV-1 response to AAV-2. It is interesting that there is a linear relationship between AAV-2 NAB and AAV-1 NAB in animals that had both.

Analysis of sera from normal human subjects was remarkable for a relative lack of NAB to either virus, with 71% of subjects scoring negative for both AAV-1 and AAV-2. Only 1 of 77 individuals was positive for AAV-1 and not AAV-2 (the titer in this person was only twofold higher than the baseline), whereas 6 individuals (8%) were positive for AAV-2 but not AAV-1. Overall, more patients had NAB to AAV-2 (27%) than to AAV-1 (20%), and the overall titers against AAV-2 appeared to be higher. These data support the hypothesis that AAV-2 infections are present in the human population, although independent primary infections with AAV-1 cannot be ruled out. The lack of monospecific antibodies to AAV-1 argues against this virus infecting humans to any appreciable degree.

Applications of AAV-1 vectors in animal models of gene therapy. AAV-1 could have a role in human gene therapy in situations in which AAV-2 is rendered ineffective because of preexisting NAB, such as in patients with a prior history of AAV-2 infection or those who received gene therapy with an AAV-2 vector. This hypothesis presumes that there is little cross-neutralization *in vivo* between AAV-1 and AAV-2. Experiments were designed to test this hypothesis in the context of AAV-mediated gene transfer to murine liver and muscle. The basic paradigm is to inject AAV-1 or AAV-2 expressing human α 1-antitrypsin and to follow up with a second vector of the same or different serotype and expressing murine erythropoietin. The efficiency of gene transfer is quantitatively measured by ELISA analysis of erythropoietin and α 1-antitrypsin in blood.

An analysis of vector readministration in muscle is presented in Fig. 7. Administration of AAV-2 vectors led to high levels of AAV-2 NAB that blocked the readministration of AAV-2 *in vivo* (group 1). The NAB in these animals did not neutralize AAV-1 *in vitro*, although there was a modest diminution (i.e., fivefold) in the transduction of an AAV-1 vector *in vivo* (group 5) to a level below that achieved in naive animals (group 4). Similarly, AAV-1 vectors resulted in high levels of AAV-1 NAB that blocked AAV-1 readministration *in vivo* (group 2) but did not interfere with AAV-2 vector transduction *in vivo* (group 6), compared to the results for naive animals that received AAV-2 alone (group 3).

Similar studies performed with a model of liver-directed gene transfer, with vectors being injected into the portal circulation via the spleen, gave mixed results (Fig. 8). Previous exposure to AAV-1 does not interfere with AAV-2 gene transfer. However, the opposite is not true. Initial treatment with AAV-2 elicited high levels of NAB to AAV-2 and low levels of NAB to AAV-1 sufficient to suppress the gene transfer of AAV-1 20-fold.

DISCUSSION

AAV vectors have been based primarily on serotype 2, a human-derived parvovirus (19, 23). The early availability of an infectious clone of AAV-2 stimulated work on the development of replication-defective vectors. The utility of AAV-2-based vectors for achieving long-term, stable, and safe gene transfer has been demonstrated with small-animal models; a

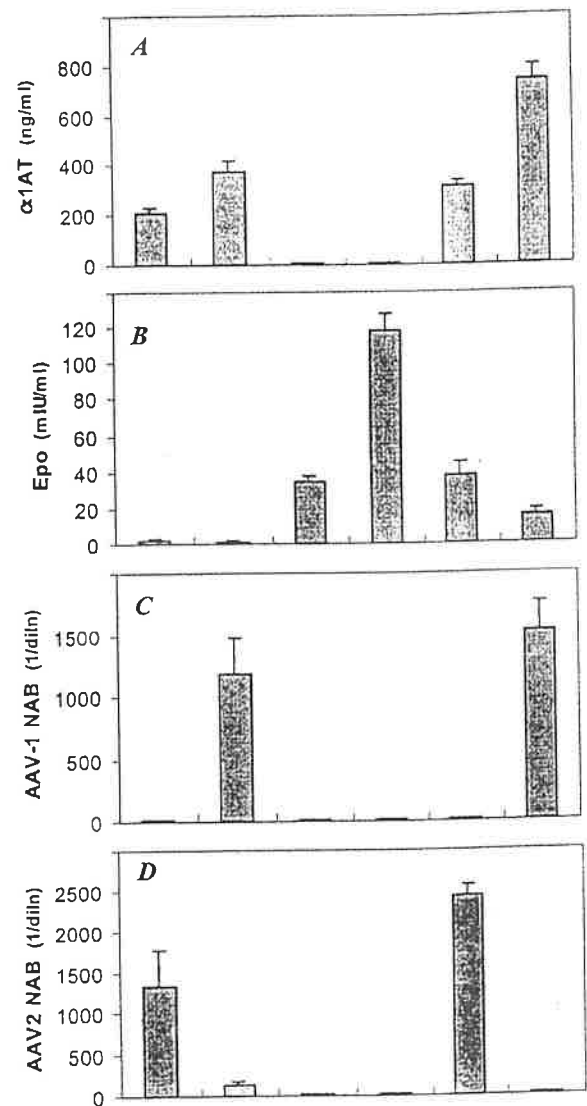


FIG. 7. Readministration of AAV vectors in muscle. C57BL/6 mice were evaluated for AAV-mediated gene transfer following introduction into naive animals as well as 30 days following the first vector administration. The different experimental groups are shown below the graphs. In each case, vector 1 expressed α 1-antitrypsin (α 1AT) from a cytomegalovirus-enhanced β -actin promoter, while vector 2 expressed murine erythropoietin (EPO) from the same promoter. Group 1, AAV-2 followed by AAV-2; group 2, AAV-1 followed by AAV-1; group 3, phosphate-buffered saline (PBS) followed by AAV-2; group 4, PBS followed by AAV-1; group 5, AAV-2 followed by AAV-1; group 6, AAV-1 followed by AAV-2. (A) Serum α 1-antitrypsin 30 days after vector 1 administration. (B) Serum erythropoietin (Epo) measured by an ELISA 30 days after vector 2 administration. (C) Reciprocal dilution (diln) of NAB to AAV-1 at day 30. (D) Reciprocal dilution of NAB to AAV-2 at day 30.

number of studies with large animals, including humans, are either planned or under way (9, 10, 21, 27). The tropism of replication-defective AAV-2 for efficient *in vivo* gene therapy is narrow.

The goal of this study was to isolate an infectious clone of AAV-1, which would be sequenced and used to develop a recombinant vector. The hypothesis was that AAV-1 vectors

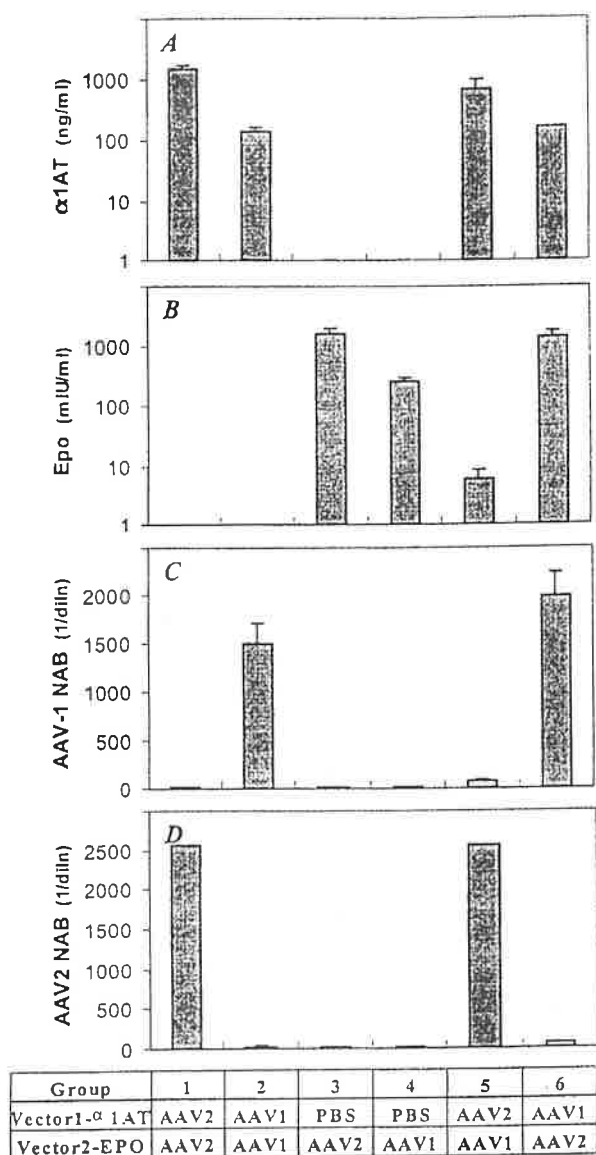


FIG. 8. Readministration of AAV vectors in the liver. These experiments are identical to those described in the legend to Fig. 7, except that animals received vectors in the portal circulation to target the liver.

would have distinct biological profiles with advantages over AAV-2 vectors in some applications. Furthermore, AAV-1 and AAV-2 vectors should not cross-neutralize, providing important advantages when used in the context of preexisting immunity due to a natural infection or prior gene therapy treatment.

An infectious clone of AAV-1 was generated from a replication intermediate. Sequencing of the AAV-1 genome confirmed the presence of the structural components that characterize the AAV family. There is approximately 80% homology in nucleotide sequence between AAV-1 and AAV-2. Delineating the entire sequence of AAV-1 has clarified a potential etiology of AAV-6, which was previously isolated as a contaminant in a human adenovirus preparation. AAV-6 appears to be a hybrid of AAV-1 and AAV-2 formed by homologous recombination between highly conserved regions spanning 452 to 552 nucleotides. There has been minimal divergence from the precursor genomes since this homologous event occurred.

The implications of this observation for gene therapy are unclear. The homologous sequences at the site of potential recombination are located at the 5' end of the Rep gene, which is deleted in the AAV vectors. Recombination of this kind may lead to additional diversity within the AAV family, although it is not clear if AAV-6 resulted from recombination *in vitro* or occurred *in vivo* during naturally acquired infections.

The initial strategy for exploiting the biology of AAV-1 for vectors was to create actual chimeras in which the vector genome was formed from AAV-2 ITRs, whereas the Rep and Cap sequences used to generate the virions were derived from AAV-1. Our goal was to retain the favorable biology of AAV-2 in terms of efficient and stable transduction while modifying the entry pathways and potentially diverting humoral immune responses (8, 31). Analysis of these vectors with murine models of liver and lung gene therapy demonstrated some important differences. When normalized for equivalent doses of administered vector genomes, AAV-2 performs better in liver than in muscle, whereas the opposite is true of AAV-1. This finding presumably is due to differences in the efficiency of entry, although postentry events that differ between the two vectors cannot be ruled out.

Another important advantage of alternative serotypes is to avoid neutralization due to preexisting humoral immunity. The existence of monospecific NAB to AAV-1 in nonhuman primates and to AAV-2 in humans supports the same serospecificity of these AAV. It should be noted, however, that seropositivity to AAV-1 and AAV-2, which is not neutralizing, is detected at higher frequencies with Western assays or ELISAs (data not shown). Administration of the vectors into murine muscle demonstrated the predicted result, in that previous treatment with AAV-2 blocks the subsequent administration of AAV-2 while having only a modest impact on the efficiency of AAV-1 treatment. The opposite is also true, in that AAV-1 administered to muscle blocks a second treatment with AAV-1 but does not interfere with gene transfer mediated by AAV-2. The murine model of AAV gene transfer to liver yielded mixed results. Clearly, AAV-2 and AAV-1 completely blocked readministration of the same serotype. Initial treatment with AAV-1 did not affect a follow-up administration of AAV-2, although the opposite was not true. Specifically, initial treatment with AAV-2 diminished retreatment with AAV-1 approximately 20-fold, a value which will likely be significant in therapeutic applications. We conclude that AAV-2 generates cross-neutralizing antibodies to AAV-1 that have an impact on AAV-1-mediated gene transfer. This effect is more significant in the context of intravascular administration. There is no evidence that antibodies to AAV-1 cross-neutralize AAV-2-mediated gene transfer to muscle or liver.

One should consider these results in the context of human applications of AAV vectors. The majority of humans do not have NAB to either AAV-1 or AAV-2. In fact, the absence of monospecific antibodies to AAV-1 in humans, together with the fact that it was isolated from monkeys, argues against it being a human virus. What is the relevance of neutralizing activity to AAV-1 in 20% of humans? Does this value reflect primary infection or cross-neutralization in the context of infection with wild-type AAV-2? Will this neutralizing activity have an impact on the efficiency of AAV-1 uptake? The data presented here suggest that the AAV-1 vector would be the preferred initial vector for muscle-directed gene therapy. It is more efficient than AAV-2 and does not preclude follow-up treatment with AAV-2. The situation for the liver is less clear. The AAV-1 vector is less efficient in the context of naive animals, although initial treatment with AAV-2 partially inhibits a second administration of AAV-1.

The impact of humoral immune responses to AAV vectors on human applications of gene therapy is an important consideration for chronic diseases. It appears that AAV expression will persist for a prolonged period of time in most target organs. What is less clear is how long the humoral immune response will persist. The use of vectors based on different serotypes should allow at least two treatments. A requirement for more frequent readministration of vectors over short time periods may require other strategies which blunt the initial humoral immune response (12, 16).

ACKNOWLEDGMENTS

We thank the Vector, Cell Morphology and Immunology Cores of the Institute for Human Gene Therapy and Wei Cao, Marcia Houston-Leslie, Rosalind Barr, Ruth Qian, George Qian, and Sarah Ehlen-Haecker for technical support.

This work was supported by grants from the NIH (P30 DK47757-06 and P01 HD32649-04), the Muscular Dystrophy Association, and Genovo, Inc., a company that J. M. Wilson founded and has equity in.

REFERENCES

- Balague, C., M. Kalla, and W. W. Zhang. 1997. Adeno-associated virus Rep78 protein and terminal repeats enhance integration of DNA sequences into the cellular genome. *J. Virol.* 71:3299-3306.
- Bantel-Schaal, U., and H. zur Hausen. 1984. Characterization of the DNA of a defective human parvovirus isolated from a genital site. *Virology* 134:52-63.
- Berns, K. I. 1995. Parvoviridae: the viruses and their replication, p. 1007-1041. In B. N. Fields, D. M. Knipe, and P. M. Howley (ed.), *Fundamental virology*, 3rd ed., vol. 2. Lippincott-Raven Publishers, Philadelphia, Pa.
- Chiorini, J. A., L. Yang, Y. Liu, B. Safer, and R. M. Kotin. 1997. Cloning of adeno-associated virus type 4 (AAV4) and generation of recombinant AAV4 particles. *J. Virol.* 71:6823-6833.
- Clark, K. R., F. Voulgaropoulou, D. M. Fraley, and P. R. Johnson. 1995. Cell lines for the production of recombinant adeno-associated virus. *Hum. Gene Ther.* 6:1329-1341.
- Clark, K. R., F. Voulgaropoulou, and P. R. Johnson. 1996. A stable cell line carrying adenovirus-inducible rep and cap genes allows for infectivity titration of adeno-associated virus vectors. *Gene Ther.* 3:1124-1132.
- Fisher, K. J., G. P. Gao, M. D. Weitzman, R. DeMatteo, J. F. Burda, and J. M. Wilson. 1996. Transduction with recombinant adeno-associated virus for gene therapy is limited by leading-strand synthesis. *J. Virol.* 70:520-532.
- Fisher, K. J., K. Jooss, J. Alston, Y. Yang, S. E. Haecker, K. High, R. Pathak, S. E. Raper, and J. M. Wilson. 1997. Recombinant adeno-associated virus for muscle directed gene therapy. *Nat. Med.* 3:306-312.
- Flotte, T., B. Carter, C. Conrad, W. Guggino, T. Reynolds, B. Rosenstein, G. Taylor, S. Walden, and R. Wetzel. 1996. A phase I study of an adeno-associated virus-CFTR gene vector in adult CF patients with mild lung disease. *Hum. Gene Ther.* 7:1145-1159.
- Flotte, T. R., and B. J. Carter. 1997. In vivo gene therapy with adeno-associated virus vectors for cystic fibrosis. *Adv. Pharmacol.* 40:85-101.
- Gao, G., G. Qu, L. Z. Faust, R. K. Engdahl, W. Xiao, J. V. Hughes, P. W. Zoltick, and J. M. Wilson. 1998. High-titer adeno-associated viral vectors from a Rep/Cap cell line and hybrid shuttle virus. *Hum. Gene Ther.* 9:2353-2362.
- Halbert, C. L., T. A. Standaert, C. B. Wilson, and A. D. Miller. 1998. Successful readministration of adeno-associated virus vectors to the mouse lung requires transient immunosuppression during the initial exposure. *J. Virol.* 72:9795-9805.
- Kaplitt, M. G., P. Leone, R. J. Samulski, X. Xiao, D. W. Pfaff, K. L. O'Malley, and M. J. During. 1994. Long-term gene expression and phenotypic correction using adeno-associated virus vectors in the mammalian brain. *Nat. Genet.* 8:148-154.
- Kotin, R. M., M. Siniscalco, R. J. Samulski, X. D. Zhu, L. Hunter, C. A. Laughlin, S. McLaughlin, N. Muzyezka, M. Rocchi, and K. I. Berns. 1990. Site-specific integration by adeno-associated virus. *Proc. Natl. Acad. Sci. USA* 87:2211-2215.
- Li, J., R. J. Samulski, and X. Xiao. 1997. Role for highly regulated rep gene expression in adeno-associated virus vector production. *J. Virol.* 71:5236-5243.
- Manning, W. C., S. Zhou, M. P. Bland, J. A. Escobedo, and V. Dwarki. 1998. Transient immunosuppression allows transgene expression following readministration of adeno-associated viral vectors. *Hum. Gene Ther.* 9:477-485.
- Muramatsu, S., H. Mizukami, N. S. Young, and K. E. Brown. 1996. Nucleotide sequencing and generation of an infectious clone of adeno-associated virus 3. *Virology* 221:208-217.
- Murphy, F. A., C. M. Fauquet, M. A. Mayo, A. W. Jarvis, S. A. Ghabrial, M. D. Summers, G. P. Martelli, and D. H. L. Bishop. 1995. Classification and nomenclature of viruses: sixth report of the International Committee on Taxonomy of Viruses. *Arch. Virol.* 1995:169-175.
- Muzyezka, N. 1992. Use of adeno-associated virus as a general transduction vector for mammalian cells. *Curr. Top. Microbiol. Immunol.* 158:97-129.
- Parks, W. P., D. W. Boucher, J. L. Melnick, L. H. Taber, and M. D. Yow. 1970. Seroepidemiological and ecological studies of the adenovirus-associated satellite viruses. *J. Virol.* 2:716-722.
- Rubenstein, R. C., U. McVeigh, T. R. Flotte, W. B. Guggino, and P. L. Zeitlin. 1997. CFTR gene transduction in neonatal rabbits using an adeno-associated virus (AAV) vector. *Gene Ther.* 4:384-392.
- Rutledge, E. A., C. L. Halbert, and D. W. Russell. 1998. Infectious clones and vectors derived from adeno-associated virus (AAV) serotypes other than AAV type 2. *J. Virol.* 72:309-319.
- Samulski, R. J. 1993. Adeno-associated virus: integration at a specific chromosomal locus. *Curr. Opin. Genet. Dev.* 3:74-80.
- Samulski, R. J., K. I. Berns, M. Tan, and N. Muzyezka. 1982. Cloning of adeno-associated virus into pBR322: rescue of intact virus from the recombinant plasmid in human cells. *Proc. Natl. Acad. Sci. USA* 79:2077-2081.
- Samulski, R. J., A. Srivastava, K. I. Berns, and N. Muzyezka. 1983. Rescue of adeno-associated virus from recombinant plasmids: gene correction within the terminal repeats of AAV. *Cell* 33:135-143.
- Samulski, R. J., X. Zhu, X. Xiao, J. D. Brook, D. E. Housman, N. Epstein, and L. A. Hunter. 1991. Targeted integration of adeno-associated virus (AAV) into human chromosome 19. *EMBO J.* 10:3941-3950. (Erratum, 11:1228, 1992.)
- Snyder, R. O., C. H. Miao, G. A. Patijn, S. K. Spratt, O. Danos, D. Nagy, A. M. Gown, B. Winther, L. Meuse, L. K. Cohen, A. R. Thompson, and M. A. Kay. 1997. Persistent and therapeutic concentrations of human factor IX in mice after hepatic gene transfer of recombinant AAV vectors. *Nat. Genet.* 16:270-276.
- Song, S., M. Morgan, T. Ellis, A. Poirier, K. Chesnut, J. Wang, M. Brantly, N. Muzyezka, B. J. Byrne, M. Atkinson, and T. R. Flotte. 1998. Sustained secretion of human alpha-1-antitrypsin from murine muscle transduced with adeno-associated virus vectors. *Proc. Natl. Acad. Sci. USA* 95:14384-14388.
- Surosky, R. T., M. Urabe, S. G. Godwin, S. A. McQuiston, G. J. Kurtzman, K. Ozawa, and G. Natsoulis. 1997. Adeno-associated virus Rep proteins target DNA sequences to a unique locus in the human genome. *J. Virol.* 71:7951-7959.
- Xiao, W., S. C. Berta, M. M. Lu, A. D. Moscioni, J. Tazelaar, and J. M. Wilson. 1998. Adeno-associated virus as a vector for liver-directed gene therapy. *J. Virol.* 72:10222-10226.
- Xiao, W., and J. M. Wilson. Unpublished data.
- Xiao, X., J. Li, and R. J. Samulski. 1996. Efficient long-term gene transfer into muscle tissue of immunocompetent mice by adeno-associated virus vector. *J. Virol.* 70:8098-8108.
- Xiao, X., J. Li, and R. J. Samulski. 1998. Production of high-titer recombinant adeno-associated virus vectors in the absence of helper adenovirus. *J. Virol.* 72:2224-2232.

Protein

GenPept

capsid protein [Adeno-associated virus - 1]

NCBI Reference Sequence: NP_049542.1

[Identical Proteins](#) [FASTA](#) [Graphics](#)[Go to:](#)

LOCUS NP_049542 736 aa linear VRL 11-MAR-2010
 DEFINITION capsid protein [Adeno-associated virus - 1].
 ACCESSION NP_049542
 VERSION NP_049542.1
 DBLINK BioProject: [PRJNA15323](#)
 DBSOURCE REFSEQ: accession [NC_002077.1](#)
 KEYWORDS RefSeq.
 SOURCE Adeno-associated virus - 1
 ORGANISM [Adeno-associated virus - 1](#)
 Viruses; ssDNA viruses; Parvoviridae; Parvovirinae;
 Dependoparvovirus.
 REFERENCE 1 (residues 1 to 736)
 AUTHORS Xiao,W., Chirmule,N., Berta,S.C., McCullough,B., Gao,G. and
 Wilson,J.M.
 TITLE Gene therapy vectors based on adeno-associated virus type 1
 JOURNAL J. Virol. 73 (5), 3994-4003 (1999)
 PUBMED [10196295](#)
 REFERENCE 2 (residues 1 to 736)
 CONSRM NCBI Genome Project
 TITLE Direct Submission
 JOURNAL Submitted (01-AUG-2000) National Center for Biotechnology
 Information, NIH, Bethesda, MD 20894, USA
 REFERENCE 3 (residues 1 to 736)
 AUTHORS Xiao,W. and Wilson,J.M.
 TITLE Direct Submission
 JOURNAL Submitted (05-MAY-1998) IHGT, University of Pennsylvania, 3601
 Spruce Street, Philadelphia, PA 19104, USA
 COMMENT PROVISIONAL REFSEQ: This record has not yet been subject to final
 NCBI review. The reference sequence is identical to [AAD27757](#).
 Method: conceptual translation.
 FEATURES
 source 1..736
 /organism="Adeno-associated virus - 1"
 /db_xref="taxon:[85106](#)"
 /note="AAV1"
 Protein 1..736
 /product="capsid protein"
 /calculated_mol_wt=81244
 Region 48..110
 /region_name="Parvo_coat_N"
 /note="Parvovirus coat protein VP1; pfam08398"
 /db_xref="CDD:[285583](#)"
 Region 220..708
 /region_name="Parvo_coat"
 /note="Parvovirus coat protein VP2; pfam00740"
 /db_xref="CDD:[279128](#)"
 CDS 1..736
 /locus_tag="AAV1gp2"
 /coded_by="NC_002077.1:2223..4433"
 /db_xref="GeneID:[1450374](#)"
 ORIGIN
 1 maadgylpdw lednlsegir ewwdlkgpaa kpkanaqqkd dgrglvlpgy kylgpfnlgl
 61 kgepvnaada aalehdkayd qqlkagdnpy lrynhadaef qerlqedtsf ggnlgravfq
 121 akkrvleplg lveegaktap gkkrpveqsp qepdsssgig ktgqqpakkr lnfgqtgdse
 181 svdpqpplge ppatpaavgp ttmasgggap madnnegadg vgnasgnwhc dstwlgdrvi
 241 ttstrtwalp tynnhlykqi ssastgasnd nhyfgystpw gyfdfnrfhc hfsprdwqrl
 301 innnwgfrrp rlnfklfnlq vkevtndgv ttianntlst vqvfsdseyq lpyvlgshaq
 361 gclppfpadv fmipqygylt lnngsqavgr ssfycleyfp sqmlrtgnnf tfsytfeevp
 421 fhssyahsqsl dlrlnmnlid qyllylnrtq nqsgsaqnkd llfsrgspag msvqpknlwp
 481 gpcyrqqrsv ktktdnnnsn ftwtgaskyn lngresiinp gtamashkdd edkffpmsgv
 541 mifgkesaga sntaldnvmi tdeeeikatk pvaterfgtv avnfqssstd patgdvhamg
 601 alpgmvwqdr dylqlqpwa kiphtdghfh psplmggfgl knpppqilik ntpvpanppa
 661 efsatkfksf itqystgqvs veiewelqke nskrwnpeqv ytsnyaksan vdfvtvdnngl
 721 yteppigtr yltrpl
 //

Directed Evolution of Adeno-associated Virus for Enhanced Gene Delivery and Gene Targeting in Human Pluripotent Stem Cells

Prashanth Asuri¹⁻³, Melissa A Bartel¹⁻³, Tandis Vazin¹⁻³, Jae-Hyung Jang¹⁻⁴, Tiffany B Wong⁵ and David V Schaffer¹⁻³

¹Department of Chemical and Biomolecular Engineering, University of California, Berkeley, California, USA; ²Department of Bioengineering, University of California, Berkeley, California, USA; ³The Helen Wills Neuroscience Institute, University of California, Berkeley, California, USA; ⁴Department of Chemical and Biomolecular Engineering, Yonsei University, Seoul, Korea; ⁵Department of Molecular and Cell Biology, University of California, Berkeley, California, USA

Efficient approaches for the precise genetic engineering of human pluripotent stem cells (hPSCs) can enhance both basic and applied stem cell research. Adeno-associated virus (AAV) vectors are of particular interest for their capacity to mediate efficient gene delivery to and gene targeting in various cells. However, natural AAV serotypes offer only modest transduction of human embryonic and induced pluripotent stem cells (hESCs and hiPSCs), which limits their utility for efficiently manipulating the hPSC genome. Directed evolution is a powerful means to generate viral vectors with novel capabilities, and we have applied this approach to create a novel AAV variant with high gene delivery efficiencies (~50%) to hPSCs, which are importantly accompanied by a considerable increase in gene-targeting frequencies, up to 0.12%. While this level is likely sufficient for numerous applications, we also show that the gene-targeting efficiency mediated by an evolved AAV variant can be further enhanced (>1%) in the presence of targeted double-stranded breaks (DSBs) generated by the co-delivery of artificial zinc finger nucleases (ZFNs). Thus, this study demonstrates that under appropriate selective pressures, AAV vectors can be created to mediate efficient gene targeting in hPSCs, alone or in the presence of ZFN-mediated double-stranded DNA breaks.

Received 26 May 2011; accepted 27 October 2011; published online 22 November 2011. doi:10.1038/mt.2011.255

INTRODUCTION

The capacity to mediate high efficiency gene delivery to human pluripotent stem cells (hPSCs) has numerous applications, ranging from the study of specific genes in stem cell self-renewal and differentiation to the directed differentiation of stem cells into specific lineages for therapeutic application. Furthermore, precise manipulation of the human stem cell genome using gene-targeting techniques that exploit the natural ability of cells to

perform homologous recombination (HR) has broad applications and implications, including safe harbor integration of genes for basic or therapeutic application, creating *in vitro* models for investigating human development and disease, and high-throughput drug discovery and toxicity studies.^{1,2} However, while gene delivery and gene targeting are well established for various mammalian somatic cells,³ a readily generalized approach for efficient gene expression and gene targeting in hPSCs requires further development.

Current methods to deliver genes to hPSCs range from viral vectors to plasmid-based transient gene expression. Lentiviral vectors—which are highly efficient and result in long-term gene expression—have been extensively employed in numerous studies in human stem cells.⁴ However, transient expression is desirable in some cases, such as for the temporary overexpression of regulatory signals to manipulate stem cell fate decisions.^{5,6} Also, vector integration into the genome can risk insertional mutagenesis, a potential concern for downstream clinical application.⁴ As an alternative, electroporation can be used to achieve transient gene expression, though this method can suffer from low transfection efficiencies and moderate toxicity in human stem cells.⁷ In addition to gene delivery, there is a need to develop efficient gene-targeting methods that rely on HR to introduce permanent and sequence-specific genome modifications in hPSCs. Several impressive studies have demonstrated successful gene targeting in hPSCs, though the initial rates reported using conventional methodologies are low (10^{-7} – 10^{-5} correctly targeted cells for every original cell in the population), which necessitates the use of positive and negative selection to improve the overall efficiency and specificity of the process.^{8,9}

Recently, several approaches have been utilized to improve gene targeting in hPSCs, including the introduction of double-stranded breaks (DSBs) into the cellular genome by engineered nucleases.^{10–14} Such breaks stimulate the cellular DNA repair machinery and thereby greatly enhance the rate of homologous recombination with a donor DNA.¹⁵ For example, Zou *et al.* found that cotransfection of plasmids containing the donor DNA and zinc finger nucleases (ZFNs) significantly increased gene targeting in hPSCs up to a frequency of 0.24% (~1 targeted cell in

The first two authors contributed equally to this work.

Correspondence: David V Schaffer, Department of Chemical and Biomolecular Engineering, University of California, Berkeley, 274 Stanley Hall, Berkeley, California 94720-1462, USA. E-mail: schaffer@berkeley.edu

415 original cells), as compared to less than 1 targeted cell per 10^6 original cells in the absence of the ZFNs.¹³ More recently, a new approach has been described to engineer DNA-binding specificities based on transcription activator–like effector (TALE) proteins from *Xanthomonas* plant pathogens and artificial restriction enzymes generated by fusing TALEs to the catalytic domain of *FokI* was used to generate discrete edits or small deletions within endogenous human genes at efficiencies of up to 25%.¹⁶ However, while the utilization of ZFNs and transcription activator–like effector nucleases (TALENs) to enhance gene-targeting efficiencies in hPSCs is highly promising, the approach entails custom engineering of a new nuclease for each new target locus, which requires labor, time, and resources.

Adeno-associated virus (AAV) is a nonpathogenic, nonenveloped virus containing a 4.7 kb single-stranded DNA genome that encodes the structural proteins of the viral capsid (encoded by the *cap* gene) and the nonstructural proteins necessary for viral replication and assembly (encoded by the *rep* gene), flanked by short inverted terminal repeats.¹⁷ Recombinant versions of AAV can be created by inserting a sequence of interest in place of *rep* and *cap*, and the resulting recombinant vectors can efficiently deliver a transgene and safely mediate long-term gene expression in dividing and nondividing cells of numerous tissues.¹⁷ Such AAV-based vectors have proven safe, efficient, and recently very effective for clinical application.^{18,19}

An interesting property of AAV is that, as demonstrated by Russell and colleagues, AAV vector genomes carrying gene-targeting constructs can mediate HR with target loci in a cellular genome at efficiencies 10^3 – 10^4 -fold higher than plasmid constructs.³ The AAV genome's inverted terminal repeats, which apparently mediate its entry into the RAD51/RAD54 component of the cellular HR pathway, play a central role in this property.²⁰ AAV-mediated gene targeting has been successfully applied to cells that AAV can effectively transduce, but naturally occurring AAV variants are typically inefficient at infecting a number of stem cell types, particularly human embryonic stem cells (hESCs).^{21,22}

We have implemented directed evolution—a rapid, high-throughput selection approach to create and isolate novel mutants from millions of genetic variants—to rapidly engineer AAV variants with desired gene delivery properties in the absence of the extensive mechanistic information typically required for rational biomolecular design. Recent work highlights the ability to apply directed evolution to create AAV mutants with altered receptor binding, neutralizing antibody-evasion properties, and altered cell tropism *in vitro* and *in vivo*.^{23–29} Likewise, we have recently demonstrated that directed evolution can yield novel AAV variants that enhance gene delivery and gene targeting in neural stem cells.²¹ Here, we implement this approach to create new AAV mutants with the enhanced capacity to infect and, subsequently, increase the efficiency of AAV-mediated gene targeting in hPSCs, both in the presence and absence of ZFN-mediated double-stranded DNA breaks.

RESULTS

Evaluation of wild-type serotypes

There are numerous naturally occurring AAV variants and serotypes, each of which has different protein capsids and thus

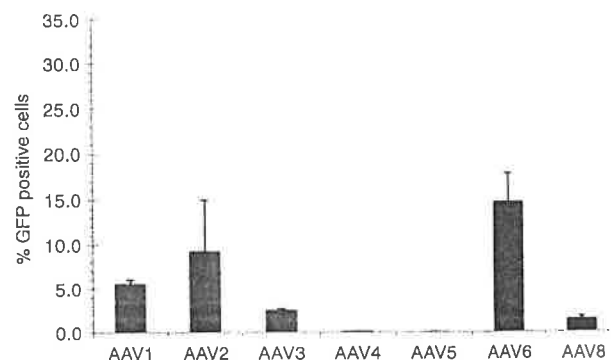


Figure 1 Gene expression in hESCs mediated by various AAV serotypes. HSF-6 cells were infected with numerous natural AAV serotypes at a MOI of 10,000. Transduction efficiency was assessed as the percentage of GFP positive cells measured by flow cytometry 48 hours postinfection. Error bars indicate standard deviation ($n = 3$). AAV, adeno-associated virus; GFP, green fluorescent protein; hESC, human embryonic stem cell; MOI, multiplicity of infection.

somewhat different gene delivery properties for various cell types and tissues.³⁰ AAV serotypes 1–6 and 8 were tested to assess their potential for hESC gene delivery. Initially, HSF-6 hESCs were infected with AAV vectors carrying green fluorescent protein (GFP) complementary DNA at a multiplicity of infection (MOI) of 10,000, and 48 hours postinfection, flow cytometry was used to determine the percentage of cells that express GFP and were thus infected by AAV. The most efficient was AAV6 (Figure 1); however, its low delivery efficiency ($14.75 \pm 3.08\%$ GFP positive cells) places significant limitations on the fraction of cells that express a transgene or could undergo gene targeting. These results are consistent with earlier reports showing that natural AAV serotypes are typically inefficient in gene delivery to human stem cells,^{31,32} thus demonstrating the need to find, or evolve, an AAV variant capable of more efficient transduction.

AAV library generation and selection through directed evolution

The AAV capsid proteins, encoded by the *cap* gene, determine the virus' ability to infect cells through their initial binding to various cell-surface receptors, intracellular trafficking, and entry into the nucleus. Directed evolution is a high-throughput approach that involves the creation and functional selection of libraries of genetic mutants to isolate novel variants with desirable properties. We implemented a directed evolution strategy to create AAV variants capable of mediating efficient gene delivery to hESCs (Figure 2a). Briefly, hESCs were infected with pools of virus variants created using several different techniques: a library in which AAV2 and AAV6 *cap* were subjected to error-prone PCR, a "shuffled" library composed of random *cap* chimeras of seven parent AAV serotypes,²⁴ an AAV2 *cap* library with random 7-mer inserts,³³ and AAV2 *cap* with substituted loop regions.²⁵ Following infection with AAV libraries (Figure 2a, step 4) and amplification of the infectious AAV variants through adenovirus superinfection—as AAV requires a helper virus such as adenovirus to induce replication—(step 5), the resulting titre of AAV rescued from each library condition was quantified and compared to titres of recovered wild-type AAV2 as a metric for relative success of

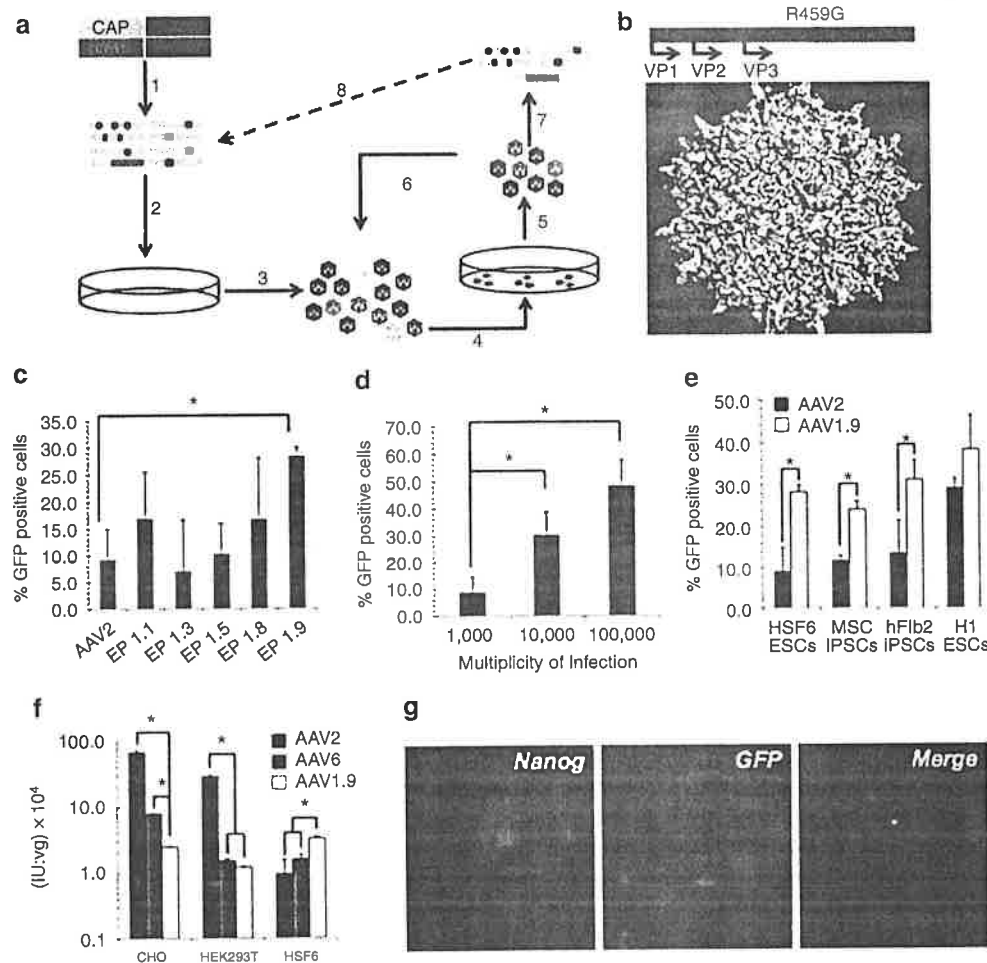


Figure 2 Directed evolution of AAV and hPSC transduction by AAV 1.9. **(a)** Schematic representation of directed evolution. 1) A viral library is created by mutating the *cap* gene. 2) Viruses are packaged in HEK293T cells using plasmid transfection, such that each particle is composed of mutant capsid surrounding the *cap* gene encoding that protein capsid. 3) Viruses are harvested from 293 cells and purified. 4) The viral library is introduced to the HSF-6 hESCs *in vitro*. 5) Successful viruses are amplified and recovered using adenovirus rescue. 6) Successful clones are enriched through repeated selections. 7) Isolated viral DNA reveals selected *cap* genes. 8) Selected *cap* genes are mutated again to serve as a new starting point for selection. **(b)** Molecular model of the full AAV2 capsid, based on the solved structure,⁵¹ shows the location of the R459G mutation (blue) on the surface of the capsid (VP3 region), near the threefold axis of symmetry and residues known to be important for heparin and FGF receptor binding (pink). **(c)** AAV-mediated gene expression in hESCs. HSF-6 cells were infected with selected mutants at a MOI of 10,000. Transgene expression was assessed as the percentage of GFP positive cells measured by flow cytometry 48 hours postinfection. Error bars indicate the standard deviation ($n = 3$), $*P < 0.01$. **(d)** Elevating the MOI increases transduction. HSF-6 cells were infected with AAV1.9 at MOIs of 1,000, 10,000, and 100,000. Transgene expression was assessed as the percentage of GFP positive cells measured by flow cytometry 48 hours postinfection. Error bars indicate standard deviation ($n = 3$), $*P < 0.01$. **(e)** AAV1.9-mediated gene expression in hPSCs. HSF-6 hESCs, human dermal fibroblast-derived hiPSCs, human MSC-derived hiPSCs, and H1 hESCs were infected at a MOI of 10,000. Transgene expression was assessed as the percentage of GFP positive cells measured by flow cytometry 48 hours postinfection. Error bars indicate the standard deviation ($n = 3$), $*P < 0.01$. **(f)** *In vitro* analysis of AAV1.9 tropism. Variant AAV1.9 (white), selected on hESCs, is less infectious on AAV-permissive cell types (HEK293T, CHO) than recombinant AAV2 (black) and AAV6 (gray), but more infectious on HSF-6 hESCs. Error bars indicate standard deviation ($n = 3$), $*P < 0.01$. **(g)** HSF-6 cells infected by AAV1.9 maintain pluripotency marker expression. HSF-6 cells infected with AAV1.9 encoding GFP at an MOI of 100,000 were fixed and immunostained 1 week after infection for the presence of GFP (green) and the pluripotency marker Nanog (red). AAV, adeno-associated virus; FGF, fibroblast growth factor; GFP, green fluorescent protein; hESC, human embryonic stem cell; hiPSC, human induced pluripotent stem cell; MSC, mesenchymal stem cell.

the selection. For each selection step, viral pools from the library produced higher viral titres than wild-type AAV2 at MOIs of both 10 and 100 (data not shown). These viral pools were then used as the starting point for the subsequent selection step (step 6). After three such selection steps, the successful viral *cap* genes were isolated (step 7) and tested individually to identify variants with the most efficient gene delivery. In addition, the *cap* genes isolated after the third selection step were subjected to an additional round

of evolution, *i.e.*, additional mutagenesis (step 8) and three selection steps, to further increase the fitness of the pool.

Increased transduction efficiency of the novel evolved AAV variant in hESCs

Nine *cap* variants from each library isolated after the third selection step of the first round of evolution were fully sequenced, and the AAV capsid protein variations and the frequency with which each

clone was detected in the error-prone PCR-mutagenized AAV2 and AAV6 library are shown in Table 1. The *cap* genes of these variants were also used to package recombinant AAV carrying the GFP gene, under the control of the cytomegalovirus (CMV) promoter. These recombinant AAV variants were then used to infect hESCs, and the gene delivery efficiency of each virus was measured via flow cytometry. Of the mutants isolated from the error-prone PCR-mutagenized AAV2 and AAV6 library, the variant AAV EP 1.9 (or AAV1.9), which carries a single-point mutation (R459G) (Figure 2b), showed the highest infection efficiency at $28.25 \pm 1.68\%$ GFP positive cells, representing an approximately threefold increase over AAV2 (its parental serotype) and a two-fold increase over AAV6 (the best serotype) at an MOI of 10,000 (Figure 2c). Variants isolated from the shuffled, 7-mer insertion, and substituted loop region libraries were sequenced and analyzed as well. Variants from these libraries showed increased infection efficiency compared to AAV2, but not AAV6 (data not shown). At an MOI of 100,000, the variant AAV1.9 achieved a strong infection efficiency of $48.21 \pm 12.92\%$ GFP positive cells (Figure 2d). It should be noted that hESCs grow as colonies that are typically multilayered, making accessing and infecting all of the cells in a colony difficult. However, AAV1.9 provides an increase in gene delivery efficiency, by an increase in the number of cells infected (Figure 2c), an increase in the relative fluorescence intensity of infected cells (approximately threefold higher than AAV6), and a significantly increased number of donor DNA molecules

Table 1 Summary of variants isolated from three rounds of selection against hESCs

Clone name	Mutations	Frequency
EP 1.1	S85G, R459G	3/9
EP 1.3	R459G, V600A	1/9
EP 1.5	S85G, R447S, R459G	1/9
EP 1.8	R447S, R459G	2/9
EP 1.9	R459G	2/9

Protein sequence features and the frequency of appearance during analysis are listed for each clone isolated from the EP AAV2/AAV6 library. AAV, adeno-associated virus; hESC, human embryonic stem cell.

delivered to each cell (Supplementary Materials and Methods and Supplementary Figure S1). Interestingly, sequencing of *cap* variants isolated after an additional round of mutagenesis to the *cap* library and three additional selection steps all matched the sequence of AAV1.9 (data not shown).

AAV1.9 was isolated through multiple rounds of selection using the HSF-6 hESC line; however, the increase in gene delivery efficiency also extended to the broadly utilized H1 hESC line, as well as induced pluripotent cell lines including mesenchymal stem cell (MSC)-derived, and dermal fibroblast-derived hiPSC lines (Figure 2e).³⁴ In addition, while AAV1.9 is capable of higher gene delivery efficiency to hESCs and hiPSCs, it is less infectious towards several cell types typically permissive to AAV infection (Figure 2f), indicating that in this case directed evolution yielded a variant that efficiently and somewhat selectively transduces an ordinarily nonpermissive cell. Furthermore, the hESCs maintained pluripotency marker expression following infection, assessed via Nanog immunostaining (Figure 2g).

The R459G mutation in AAV1.9 lies close to but not within the heparin or fibroblast growth factor (FGF) receptor binding domains on the AAV capsid surface (Figure 2b), though it could conceivably modulate binding to the cell surface. To determine whether the amino acid 459 mutation alters the heparin-binding properties of the variant, AAV2, AAV6, and AAV1.9 were analyzed using heparin column chromatography, which can provide a measure of the heparin-binding affinity of AAV variants. Despite its mutation residing outside the reported heparin-binding domain,³⁵ and despite the loss of a positively charged residue, AAV1.9 displays a higher affinity for heparin compared to the parental serotype AAV2, as indicated by the higher NaCl concentration needed to elute the majority of AAV1.9 from the heparin column. (Figure 3a). However, *in vitro* transduction of CHO and pgsA (a mutant CHO cell line lacking surface glycosaminoglycans (GAGs)) cells showed that AAV1.9 had less dependence on cell-surface heparan sulfate proteoglycans (AAV2's primary receptor)³⁶ for cell transduction (Figure 3b). Furthermore, while it exhibited higher transduction levels, AAV1.9 bound to hESCs at levels similar to AAV2 (Figure 3c). Collectively, these data indicate that the

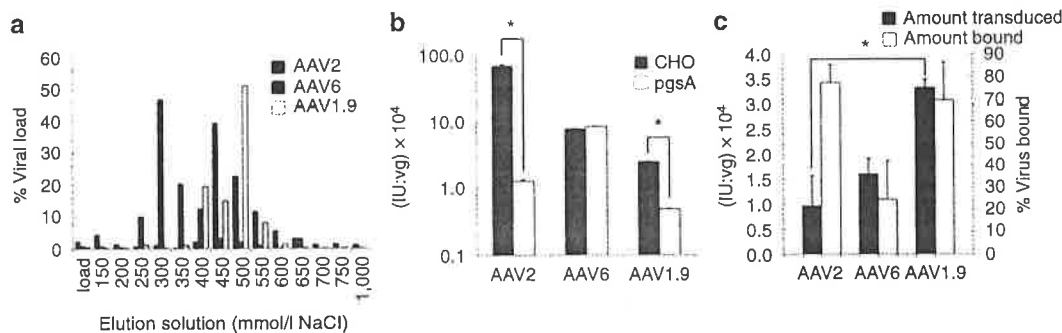


Figure 3 Mechanistic analysis of transduction by variant AAV. (a) AAV1.9 has a higher affinity for heparin than AAV2 and AAV6. The heparin affinity column chromatogram of recombinant AAV2 (black), AAV6 (gray), and AAV1.9 (white) is shown, where virus was eluted from the column using increasing concentrations of NaCl. Virus was quantified using qPCR. (b) *In vitro* characterization of AAV1.9 HSPG dependence. CHO (black) and pgsA (white) cells were transduced to demonstrate the decrease in HSPG dependence of AAV1.9 compared to AAV2. Error bars indicate standard deviation ($n = 3$), $*P < 0.01$. AAV, adeno-associated virus; hESC, human embryonic stem cell, qPCR, quantitative PCR. (c) *In vitro* characterization of AAV1.9 binding affinity. AAV1.9 and its parental serotype AAV2 bind hESCs to similar extents, yet AAV1.9 is capable of significantly higher transduction of hESCs. Error bars indicate standard deviation ($n = 3$), $*P < 0.01$.

R459G mutation may exert its effects after initial docking to the cell surface, though future studies will be needed to elucidate its mechanism of enhanced hESC infection.

Enhanced gene-targeting efficiency of the novel evolved AAV variant in hESCs

We assessed the ability of the novel variant to mediate gene targeting in hPSCs. To do so, we employed a mutated GFP-based reporter system, previously described by Zou *et al.*¹³ Briefly, GFP complementary DNA containing a 35bp insertion harboring a stop codon (GFPΔ35) was introduced into HSF-6 hESCs using a lentiviral vector (see Materials and Methods and Figure 4a). The donor construct (t37GFP) consisted of a promoter-less, nonfunctional GFP with a 37bp 5' truncation, as well as 290 nucleotides

of homology upstream and 1,619 nucleotides downstream of the 35bp insertion in GFPΔ35. The donor plasmid was packaged into recombinant AAV vectors (AAV2, AAV6, and AAV1.9), and a hESC line carrying the mutated GFP was infected with these gene-targeting constructs to test their capacity to mediate gene correction and thereby restore GFP fluorescence. The levels of gene correction achieved using recombinant AAV2 and AAV6 were below the limits of detection of flow cytometry in this assay (Figure 4b), whereas the variant AAV1.9 was capable of targeted gene correction resulting in fluorescent hESCs (0.12%). Furthermore, consistent with previous reports,³ we also show that at similar efficiencies of gene delivery, AAV vectors can mediate targeted gene correction at rates much higher than plasmid constructs (Figure 4b). Moreover, as discussed earlier, AAV1.9 is

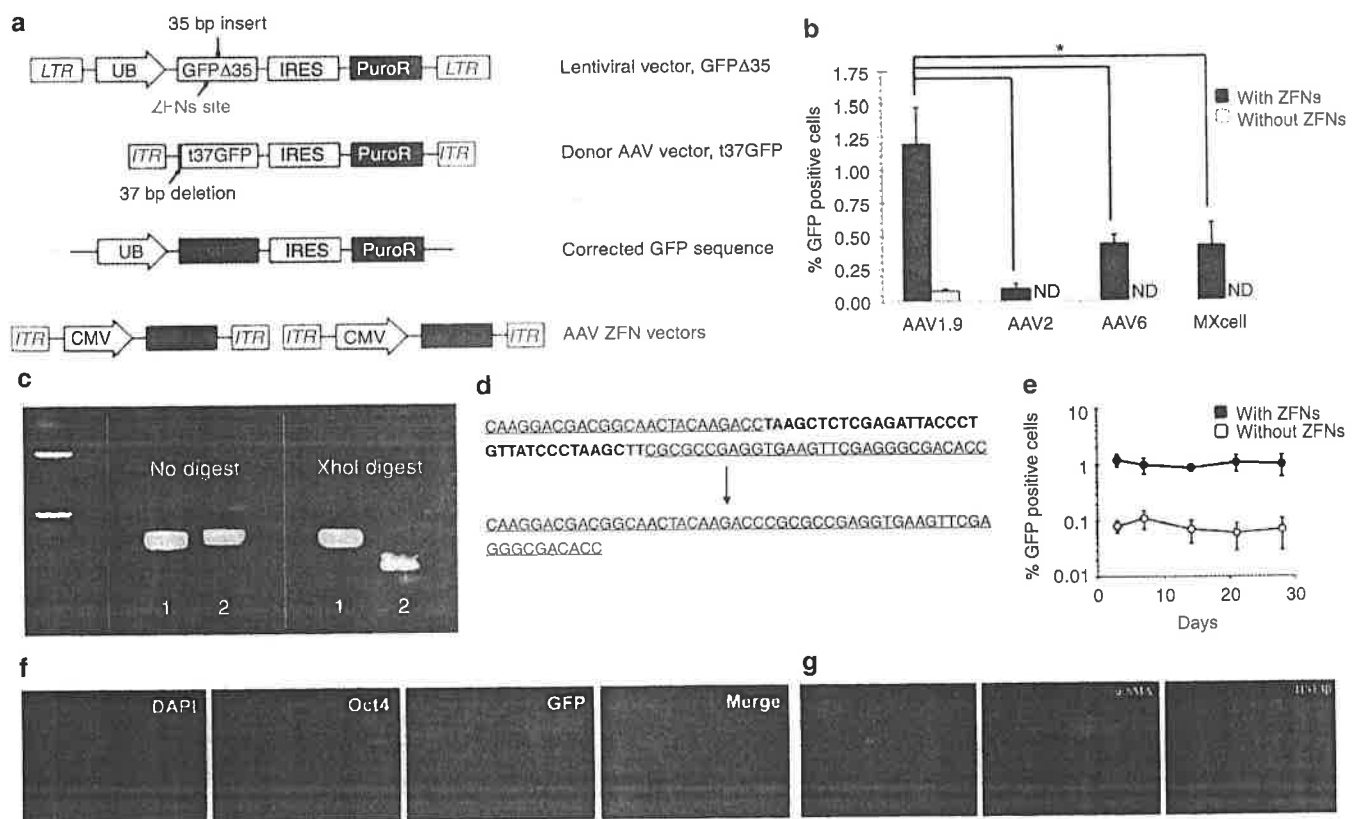


Figure 4 AAV1.9-mediated correction of a nonfunctional GFP expressed in hESCs. (a) Schematic overview depicting the targeting strategy for the GFP gene correction. The defective GFP gene—GFPΔ35, mutated by the insertion of a 35 bp fragment containing a translational stop codon¹³—was integrated into HSF-6 cells using a lentiviral vector. A targeting vector (t37GFP) containing a 5' truncated GFP coding sequence was packaged into a recombinant AAV vector lacking a promoter. Homologous recombination between donor vector and integrated defective GFPΔ35 gene would correct the 35 bp mutation and result in fluorescent hESCs. (b) Gene targeting frequencies assessed as the percentage of GFP positive cells measured via flow cytometry 72 hours postinfection. Error bars indicate standard deviation ($n = 3$), * $P < 0.01$. ND, not detectable by flow cytometry. (c,d) Representative analyses of the targeted correction of GFPΔ35 gene. Genomic DNA from the "corrected" and "uncorrected" HSF-6 cells was amplified using PCR. The 35 bp mutation within the GFPΔ35 gene contains an *Xho*I site, and therefore only the PCR products from the "uncorrected" cells (sample 2) can be digested using *Xho*I, whereas the PCR products from the corrected cells (sample 1) show the restoration of the functional GFP gene without the *Xho*I site. DNA sequencing analysis also shows AAV1.9-mediated correction of the 35 bp mutation in HSF-6 cells originally expressing the mutated GFPΔ35 gene. (e) Time course of AAV-mediated GFPΔ35 correction in HSF-6 cells. Error bars indicate standard deviation ($n = 3$). (f) Maintenance of an undifferentiated state of gene-corrected, GFP-expressing cells for 30 days postinfection. GFP positive HSF-6 cells, isolated via FACS, were cultured for 30 days, then fixed and probed for the presence of GFP (green), Oct4 (red), and DAPI (blue). (g) Pluripotency of the HSF-6 cells carrying the corrected GFP gene was further confirmed through embryoid body-mediated *in vitro* differentiation to generate derivatives of all three germ layers, as indicated by the expression of the ectodermal marker β -III tubulin, the mesodermal marker α -smooth muscle actin (α -SMA), and the endodermal marker hepatocyte nuclear factor 3 β (HNF3 β) after 24 days of differentiation (Bar = 100 μ m). AAV, adeno-associated virus; CMV, cytomegalovirus; DAPI, 4',6-diamidino-2-phenylindole; FACS, fluorescent-activated cell sorting; GFP, green fluorescent protein; hESC, human embryonic stem cell; IRES, internal ribosome entry site; ITR, inverted terminal repeat; LTR, long terminal repeat.

Table 2 Summary of gene targeting experiments and rates in various human pluripotent lines in the presence and absence of ZFN-mediated DSBs

Human pluripotent cell line	% Gene targeting without ZFNs		% Gene targeting with ZFNs	
	AAV2	AAV1.9	AAV2	AAV1.9
HSF-6 hESCs	ND	0.08 ± 0.02*	0.09 ± 0.04	1.2 ± 0.28*
H1 hESCs	0.02 ± 0.01	0.12 ± 0.03*	0.31 ± 0.07	1.6 ± 0.19*
MSC-iPSCs	ND	0.06 ± 0.01*	0.12 ± 0.04	1.3 ± 0.14*
hFib2-iPSCs	ND	0.06 ± 0.02*	0.16 ± 0.03	1.2 ± 0.19*

Values reported as mean ± SD ($n = 3$).

AAV, adeno-associated virus; hESC, human embryonic stem cell; iPSC, induced pluripotent stem cell; MSC, mesenchymal stem cell; ND, not detectable; ZFN, zinc finger nuclease.

* $P < 0.01$.

also capable of higher gene delivery efficiency to H1 hESCs, and MSC- and dermal fibroblast-derived hiPSCs (Figure 2e), and we observe a corresponding increase in gene-targeting efficiencies (0.06–0.12%) for these cell lines as well (Table 2).

Introduction of DSBs via ZFNs improves the gene-targeting efficiencies of AAV vectors

Directed evolution thus created a novel AAV variant that mediates high levels of gene targeting in hESCs and hiPSCs, and previous reports indicate that in addition to enhancing the ability of the donor DNA to mediate HR, complementary approaches such as inducing local DSBs at the target locus can further enhance gene targeting.³⁷ Recent years have witnessed the development of engineered restriction enzymes generated by fusing a DNA cleavage domain either to a zinc finger or a TAL effector DNA-binding domain as a general approach to generate site-specific double-stranded chromosomal breaks and thereby greatly stimulate homology-directed gene repair in a number of human cell lines.^{10–13,16} These findings provided rationale to explore whether stimulation of gene targeting by nuclease-mediated DSBs can be combined with an evolved AAV to further enhance gene targeting in hPSCs. We first generated AAV vectors that expressed the ZFNs from a CMV promoter to target a site 12 bp upstream of the 35 bp insertion in the GFP gene (Figure 4a).³⁸ hPSCs harboring the mutant GFP were infected with the two AAV vectors encoding a GFP-specific ZFN, as well as the AAV donor vector (t37GFP), and the gene-targeting efficiency was measured via flow cytometry to detect cells expressing the corrected GFP. In the presence of a local DSB-mediated by ZFNs, we observe a fivefold to 13-fold increase in AAV1.9-mediated gene-targeting rates (Figure 4b and Table 2), from an already relatively high efficiency of 0.06–0.12% to 1.2–1.6% across multiple hPSC lines. Also, while not as high as with AAV1.9, using other AAV serotypes to deliver the donor construct and ZFNs did yield measurable gene correction in all cases, indicating the generality of using AAV to deliver both donor and nuclease for enhanced HR (Figure 4b and Table 2). Finally, we observed similar increases in gene-targeting efficiencies when we infected human pluripotent embryonic carcinoma cells (NT2) and human embryonic kidney (HEK293T) cells harboring the mutant GFP with AAV2 vectors encoding the GFP-specific ZFN and the donor t37GFP plasmids, indicating the

general utility of combining an AAV donor with DSBs at the target locus (Supplementary Figure S2).

To verify correction of the GFPΔ35 sequence, genomic DNA from “corrected” (GFP positive cells isolated using fluorescence-activated cell sorting) and “uncorrected” HSF-6 hESCs was subjected to PCR to amplify the GFP sequence. The 35 bp insert of the mutated GFP cell line harbors a unique *Xho* I site that enabled the confirmation of GFP correction by restriction digest analysis (Figure 4c). Also, direct sequencing of the PCR fragments confirmed the correction of the mutated GFPΔ35 gene (Figure 4d). Furthermore, the GFP positive hESCs, purified via fluorescence-activated cell sorting, were expanded and monitored by flow cytometry for 30 days, and >95% of cells maintained GFP expression (Figure 4e). These cells also maintained a pluripotent state during long-term culture following targeted gene correction as indicated by the expression of the pluripotency marker Oct4 (Figure 4f) and their ability to differentiate into cells originating from the three germ layers via embryoid body formation *in vitro* (Figure 4g).

Finally, in addition to measuring targeted gene correction via HR, we also assessed the frequency of random chromosomal integration of AAV1.9 vectors. HSF-6 hESCs harboring GFPΔ35 were infected with AAV1.9 donor constructs and cultured for 14 days. The genomic DNA from the infected cells was subjected to PCR to specifically amplify a 822 bp fragment that partially spanned the AAV viral genome following the puromycin resistance gene (Supplementary Figure S3a), a region that should not participate in homologous recombination (Figure 4a) and would thus represent residual AAV sequence. This assay was calibrated by conducting PCR in parallel on naive hESC DNA samples spiked with different amounts of AAV1.9 donor plasmid, ranging from 1 copy of AAV plasmid per 400 cellular genome copies to 1 copy of AAV plasmid per 20 cellular genome copies. Based on the quantification of the samples and the standards, ~1 copy of AAV1.9 genome was present per 100 cells that were originally exposed to the vector (Supplementary Materials and Methods and Supplementary Figure S3b), consistent with earlier reports that demonstrate low risk of random genomic integration of AAV in human cells.^{37,39} To determine whether introduction of site-specific DSBs using ZFNs increases this low frequency of random integration, we coinfect HESCs carrying the mutant GFP with AAV1.9 vectors encoding the donor plasmid, as well as the GFP-specific ZFNs. We did not observe a significant increase in the levels of random integration in the presence of ZFNs (Supplementary Figure S3), which suggests that at least in this system the introduction of site-specific DSBs may not have a strong effect on nonhomologous integration frequencies of AAV vectors.^{37,39}

DISCUSSION

AAV has attracted increasing attention for its ability to safely and efficiently mediate gene targeting in a variety of cell types.^{3,30,32} However, prior studies have indicated that naturally occurring AAV serotypes are typically inefficient in transducing numerous classes of stem cells, which can limit also the application of AAV-mediated gene targeting to hPSCs. Recently, two studies have reported the natural serotype AAV3 as the most efficient serotype for human pluripotent lines.^{31,40} In the hESC and hiPSC cell lines we examined, we did not observe a higher efficiency for

AAV3 relative to other serotypes (Figure 2c and data not shown). Moreover, the reported transduction efficiency for AAV3 in the earlier studies was lower than 25% at a MOI of 200,000, and likely as a result, the maximum reported gene-targeting frequency was 1.3×10^{-5} .³¹ Collectively, these results indicate that different natural serotypes may be suited for different hPSC lines, but importantly that gene-targeting efforts in general may benefit from increasing AAV's delivery efficiency to hPSCs.

Directed evolution has proven to be a powerful approach to create AAV vectors with novel capabilities, and we successfully demonstrate that under appropriate selective pressures AAV can evolve for improved transduction efficiencies in human pluripotent stem cells. Specifically, we isolated a new AAV variant, AAV1.9, that exhibited an enhanced infection efficiency of ~48% at a MOI of 100,000. AAV1.9 harbors a single R459G mutation that lies in close proximity to both the heparin and FGF receptor binding domains of AAV2. This variant displayed a slightly higher affinity for heparin compared to AAV2 (Figure 3a), though we observed similar levels of AAV1.9 and AAV2 binding to hESCs (Figure 3c). In their investigation of AAV2's heparin-binding domain, Opie *et al.* found that a R459A AAV2 mutant was incapable of infecting HeLa cells, despite heparin-binding levels essentially the same as AAV2.⁴¹ The authors hypothesized that the R459A mutation may result in a defect in a later stage of viral infection.⁴¹ This finding, in conjunction with our observations that similar percentages of AAV2 and AAV1.9 bind to hESCs, suggests that the R459G mutation may increase hESC gene delivery by modulating viral infection at a point following initial cell-surface docking, for example AAV structural rearrangement during heparan sulfate binding,⁴² subsequent binding to secondary receptors, or cellular entry. Further studies are needed to elucidate such mechanisms.

An AAV variant with enhanced gene delivery to hESCs can serve as a valuable tool for numerous applications. For example, AAV could mediate the controlled overexpression of regulatory proteins, short hairpin RNAs, or microRNAs, which would enable basic investigations of the roles of key factors in stem cell self-renewal or lineage commitment, as well as therapeutic applications that rely on the efficient generation of specific cell lineages. AAV has the potential to remain episomal and thereby mediate transient expression of such factors, compared to typically constitutive expression from lentiviral vectors, which may in some cases be advantageous. Furthermore, increased gene delivery efficiencies afford the opportunity to deliver multiple cargos effectively. For example, a twofold increase in the delivery of one AAV vector would translate to an eightfold increase in the number of cells that would be transduced with three vectors carrying three different transgenes (such as the gene-targeting construct and two ZFN-encoding vectors utilized in this study). Another consequence of generating AAV variants with higher transduction efficiencies in stem cells is the corresponding increase in gene-targeting efficiencies. When AAV1.9 was used to mediate gene correction in HSF-6 hESCs, we observed gene-targeting frequencies of 8×10^{-4} , approximately tenfold higher than the reported targeting frequencies mediated by naturally occurring AAV serotypes.^{31,32} We have also observed that AAV1.9 allows for increased gene delivery and gene-targeting efficiencies to the H1 hESC line, as well as dermal fibroblast- and MSC-derived iPSCs, thus establishing the generality of the method.

It was previously shown that AAV gene targeting could be enhanced with a double-stranded DNA break introduced at the target locus by the nuclease I-SceI,^{37,39} and our results build upon this important prior work to show for the first time that AAV can function effectively in conjunction with ZFNs, which can now be engineered for genomic site specificity.^{10,12,13} In addition to ZFNs, TALE truncation variants have recently been linked to the catalytic domain of Fok I to yield a new class of nucleases that can generate discrete edits or small deletions within endogenous human genes and induce gene modification in human cells by both nonhomologous end joining and HR.¹⁶ Alternatively, the DNA recognition properties of homing endonucleases can be re-engineered to yield enzymes that recognize endogenous genes in human cells.⁴³ Such custom nucleases can be beneficial, as prior studies indicate that gene-targeting frequencies vary from locus to locus and are lower at nonexpressed genes,^{12,44} and the targeting efficiencies in this study may have benefitted from the use of a lentiviral vector—which preferentially integrates into active transcription units⁴⁵—as the target locus. Furthermore, researchers have observed that gene correction, gene knock-in, and knock-out vectors exhibit varying targeting frequencies.^{46,47} However, depending on the application, it is important to note that an appropriate AAV has the potential to mediate gene targeting at a given locus at rates that are sufficiently high to obviate the time, labor, and resources entailed in generating a custom nuclease. Furthermore, as ZFNs and other nucleases have the potential to exhibit off-target cleavage,^{48,49} the use of effective donor DNA in the absence of DSBs offers certain advantages.

In conclusion, we have used directed evolution to create a novel AAV variant with a single-point mutation that exhibited enhanced gene delivery efficiency, and subsequently gene targeting, in hPSCs. In addition, this work demonstrates that AAV-mediated gene-targeting frequencies can be further enhanced in the presence of targeted DSBs generated by the AAV vector co-delivery of ZFNs. These findings suggest that AAV vectors may find strong utility in investigations of stem cell biology and therapy, ranging from the generation of reporter cell lines to therapeutic gene correction.

MATERIALS AND METHODS

Cell lines. Cell lines were cultured at 37°C and 5% CO₂ and, unless otherwise mentioned, were obtained from the American Type Culture Collection (Manassas, VA). HEK293T and NT2 cells were cultured in Dulbecco's modified Eagle's medium supplemented with 10% fetal bovine serum (Gibco, Carlsbad, CA) and 1% penicillin/streptomycin (Invitrogen, Carlsbad, CA). CHO K1 and CHO pgsA cells were cultured in F-12K medium (ATCC) supplemented with 10% fetal bovine serum (Gibco) and 1% penicillin/streptomycin (Invitrogen). HSF-6 hESCs (UC San Francisco) were cultured on Matrigel-coated cell culture plates (BD, Franklin Lakes, NJ) in X-Vivo medium (Lonza, Norwalk, CT) supplemented with 80 ng/ml FGF-2 (PeproTech, Rocky Hill, NJ) and 0.5 ng/ml TGF-β1 (R&D Systems, Minneapolis, MN). H1 hESCs (WiCell, Madison, WI) and dermal fibroblast- and MSC-derived iPSCs (a kind gift from George Q. Daley, Children's Hospital Boston, Boston, MA) were cultured on Matrigel-coated cell culture plates (BD) in mTeSR1 maintenance medium (Stem Cell Technologies, Seattle, WA).

Library generation and viral production. Four replication-competent viral libraries were used as starting materials for selections on HSF-6 hESCs.

A random mutagenesis library was generated by subjecting *cap* genes from AAV2 and AAV6 to error-prone PCR using 5'-CATGGGAAA GGTGCCAGACG-3' and 5'-ACCATCGGCAGCCATACCTG-3' as forward and reverse primers, respectively, as previously described (EP Library).²³ Libraries consisting of AAV2 containing random 7-mer peptide inserts²³ and AAV2 containing randomized *cap* loop regions²⁵ were also utilized. Finally, a library containing shuffled DNA from the wild-type AAV1, AAV2, AAV4, AAV5, AAV6, AAV8, AAV9 *cap* genes was packaged.²⁴ Libraries were pooled for the selection steps. Following evolution, to create recombinant versions of selected viruses, the *cap* genes were inserted into the pXX2 recombinant AAV packaging plasmid using *Not* I and *Hind* III.²³ Both the replication-competent AAV library and recombinant AAV vectors expressing GFP under the control of a CMV promoter were packaged using HEK293T cells using the calcium phosphate transfection method, and the viruses were purified by iodixonal gradient centrifugation and Amicon filtration.^{23,27} DNase-resistant genomic titres were determined via quantitative PCR.²³

Library selection and evolution. One selection step is defined as hESC infection using a starting library, rescue by addition of adenovirus serotype 5 (at levels sufficient to induce a cytopathic effect 48 hours post-adenovirus infection), and harvest of successful variants. For each selection step, cells were grown on Matrigel-coated plates to eliminate the use of feeder cells, which could become infected and bias the selection towards viral variants that transduce the feeder layer. In addition, stromal cells were removed from the culture using collagenase digestion before the viral harvest to prevent the selection of viral variants that infect stromal cells instead of hESCs. One round of evolution consists of genetic diversification of the *cap* gene followed by three selection steps. Two rounds of evolution were performed, with clonal analysis (*cap* gene sequencing and hESC gene delivery assay for each selected virus) performed between each round of evolution. Following the third selection step, AAV *cap* genes were isolated from the pool of successful AAV variants and amplified via PCR. *Cap* genes were then sequenced at the University of California, Berkeley DNA sequencing facility and analyzed using Geneious software (Biomatters, Auckland, New Zealand). Three-dimensional models of the AAV2 capsid (Protein Databank accession number 1LP3) were rendered in Pymol (DeLano Scientific, San Carlos, CA).

hPSC transduction analysis. The human pluripotent cell lines were plated at a density of 10^5 cells/well 24 hours before infection. Cells were infected with rAAV-GFP at an MOI of 10^4 . Stromal cells were removed from the hESC and the iPSC culture using collagenase digestion before flow cytometry. The percentage of GFP positive cells was assessed 48 hours postinfection using a Beckman-Coulter Cytomics FC 500 flow cytometer (Beckman-Coulter, Brea, CA).

Immunofluorescence staining. Immunostaining was performed to visualize the expression of pluripotency markers following AAV infection and gene targeting. HSF-6 hESCs were plated on a 24-well plate and infected as described for the transduction analysis. Forty-eight hours postinfection, cells were fixed in 4% paraformaldehyde for 15 minutes, washed three times with phosphate-buffered saline (PBS), and blocked with 1% bovine serum albumin (BSA) and 0.1% Triton X-100 in PBS for 30 minutes. Cells were incubated overnight at 4°C with a mouse anti-Oct-3/4 primary antibody (1:200 dilution, Santa Cruz Biotechnology, Santa Cruz, CA) or anti-Nanog primary antibody (1:200 dilution, Santa Cruz Biotechnology). Cells were then washed three times with PBS and incubated with a secondary fluorescent-conjugated Alexa Fluor 647 goat anti-mouse antibody (1:250 dilution, Invitrogen) for 2 hours, followed by 4',6-diamidino-2-phenylindole (DAPI) for nuclear staining (Invitrogen) for 15 minutes. Cells were imaged using a Zeiss Axio Observer.A1 inverted microscope.

In vitro differentiation assay of pluripotency. Colonies of HSF-6 hESCs were isolated from stromal cells by collagenase enzymatic treatment and partially dissociated by gentle pipetting. The cells were cultivated as small clusters (5–10 cells) in suspension culture in ultra low-attachment plates

(Corning, Corning, NY) to generate embryoid bodies (EBs) in X-Vivo medium. After 6 days, the EBs were transferred to plates precoated with Matrigel and further differentiated for 24 days. Immunocytochemistry as described above was used to investigate whether the cells express markers of the three germ layers using the following primary antibodies: ectodermal: rabbit anti- β -III tubulin (1:500 dilution; Covance, Princeton, NJ); mesodermal: mouse anti- α -smooth muscle actin (1:500 dilution; Sigma, St Louis, MO); endodermal: rabbit anti-hepatocyte nuclear factor 3 β (1:500 dilution; Millipore, Billerica, MA). Fluorescence images were acquired using a Nikon TE2000E2 epifluorescence microscope.

In vitro transduction and cell binding analysis. To determine the heparan sulfate dependence of AAV2 and the selected mutant, CHO K1 and CHO pgsA cells were plated at a density of 2.5×10^4 cells per well 24 hours before infection. Cells were infected with AAV2-GFP, AAV6-GFP, or AAV1.9-GFP at MOIs ranging from 100–2,500.²⁶ The percentage of GFP positive cells was assessed 48 hours postinfection using a Beckman-Coulter Cytomics FC 500 flow cytometer.

To analyze cell-surface binding of AAV variants, cells were plated as described for the hESC transduction analysis. Stromal cells were removed from the hESC culture via collagenase digestion, and cells were incubated at 4°C for 10 minutes; 10^5 cells were incubated with rAAV-GFP (MOI of 10^4) at 4°C for 1 hour. Cells were washed twice with ice-cold PBS to remove unbound virus. DNase-resistant genomic titers were determined via quantitative PCR.

Heparin column chromatography. rAAV-GFP vectors were subjected to heparin column chromatography as previously described.²³ Briefly, 10^{10} vector genome containing AAV particles were loaded into a 1 ml HiTrap heparin column (GE Healthcare, Piscataway, NJ) previously equilibrated with 0.15 mol/l NaCl and 50 mmol/l Tris at pH = 7.5. Washes were performed using 1 ml volumes of Tris buffer containing increasing increments of 50 mmol/l NaCl, starting at 150 mmol/l NaCl and ending at 750 mmol/l NaCl, followed by a 1 mol/l NaCl wash. Genomic titers from each fraction were determined via quantitative PCR.

Generation of mutant GFP hPSC lines. An internal ribosome entry site and puromycin resistance gene cassette (IRES-PuroR) were cloned into the *Eco* RI and *Xho* I sites of pFUGW (a kind gift from David Baltimore, California Institute of Technology, Pasadena, CA) to replace the woodchuck hepatitis virus post-transcriptional regulatory element (WPRE) and yield pFUGIP. Next, a mutant GFP sequence harboring a 35bp insertion (GFP Δ 35) (a kind gift from Matthew H. Porteus, University of Texas Southwestern Medical Center, Dallas, TX) digested with *Bam* HI/*Eco* RI was inserted in place of the GFP sequence of similarly digested pFUGIP to construct pFUG Δ 35IP. The lentiviral vector carrying the mutated GFP was packaged by the calcium phosphate transient transfection method.²⁰ Briefly, 10 μ g of pFUG Δ 35IP, 5 μ g of pMDL g/p PRE, 3.5 μ g of pcDNA 3 IVS VSV-G, and 1.5 μ g of pRSV Rev were transfected into HEK 293T cell lines (>70% confluency), which were cultured on 10 cm tissue culture plates under Iscove's modified Dulbecco's medium (IMDM) supplemented by 10% fetal bovine serum (FBS) and 1% penicillin/streptomycin. The lentiviral vector was harvested and concentrated by ultracentrifugation (L8-55M Ultracentrifuge; Beckman Coulter), followed by resuspension in 100 μ l of PBS with 20% sucrose. Stable hPSC lines expressing the mutant GFP was generated by infection with the lentiviral vector, followed by puromycin selection (1 μ g/ml) for 1 week. GFP fluorescence was not detected in the GFP Δ 35 hPSCs, as confirmed by flow cytometry.

Gene targeting assay. The donor plasmid was generated by subcloning a truncated GFP gene missing the first 37 base pairs (t37GFP), internal ribosome entry site, and puromycin resistance gene cassette into an AAV vector plasmid (pAAV CMV GFP SN), such that the resulting targeting vector lacked a promoter. The ZFN expression constructs, driven by a CMV promoter, were generated by subcloning gene cassettes

containing the engineered zinc fingers targeting GFP (also a gift from Matthew H. Porteus) into pAAV CMV GFP SN (detailed cloning steps for the donor and ZFN expressing constructs available upon request). The donor construct and the ZFN expressing constructs were then packaged into the AAV2, AAV6, and AAV1.9 capsids. All viral vectors were harvested and purified as described earlier.

For gene-targeting experiments mediated by AAV, hPSCs were seeded onto 12-well tissue culture plates at a density of 10^5 cells/well 24 hours before AAV infection. The cells were infected with the AAV targeting vectors at an MOI of 10^5 . For targeting experiments in the presence of ZFNs, the ZFN expressing vectors were added at an MOI of 10^5 to the cells in addition to the AAV targeting vectors. For gene-targeting experiments mediated by naked plasmid constructs, Rho kinase inhibitor (ROCK inhibitor, Y-27632; CalBioChem, San Diego, CA) was added to hPSC media 24 hours before electroporation. Cells were harvested using collagenase digestion, resuspended at 10^6 cells/ml in 100 μ l Bio-Rad buffer (GPEB 2) containing 1 μ g each of the donor plasmid and ZFN expression plasmids, and electroporated using the Bio-Rad Gene Pulser MXcell System (250 V, 500 μ F, 1000 ohms; Bio-Rad, Hercules, CA). The cells were replated on Matrigel-coated plates in hPSC culture medium supplemented with ROCK inhibitor. The medium was replaced after 1 day. The percentage of GFP positive cells was assessed 48 hours postinfection using flow cytometry as described earlier.

Cell sorting and sequencing. Cells infected with AAV1.9 were cultured for 30 days and sorted at the UC Berkeley Cancer Center with a BD Influx Sorter to isolate GFP positive cells. For DNA sequencing analysis, cellular genomic DNA was extracted with the QIAamp DNA Micro Kit (Qiagen, Valencia, CA) and amplified using 5'-CCACCTCGTGACCACCCTG-3' and 5'-CGGCCATGATATAGACGTTGTGGC-3' primers. The resulting PCR products were cloned using the StrataClone™ PCR cloning kit (Stratagene, La Jolla, CA), and individual clones were sequenced to confirm the gene correction.

SUPPLEMENTARY MATERIAL

Figure S1. AAV1.9 increases the number of donor molecules per cell.

Figure S2. AAV2-mediated correction of a nonfunctional GFP expressed in HEK293Ts and NT2s.

Figure S3. Representative analysis of residual AAV1.9 genomes.

Materials and Methods.

ACKNOWLEDGMENTS

This work was supported by CIRM Award RT1-01021 and an NSF Graduate Fellowship (M.A.B.). We thank Matthew H. Porteus (UT Southwestern) and Linzhao Cheng (JHU Medicine) for the mutated and donor GFP and GFP-ZFN plasmid constructs, and George Q. Daley (Children's Hospital Boston) for dermal fibroblast- and MSC-derived iPS cells. The authors declared no conflict of interest.

REFERENCES

- Conrad, C, Gupta, R, Mohan, H, Niess, H, Bruns, CJ, Kopp, R *et al.* (2007). Genetically engineered stem cells for therapeutic gene delivery. *Curr Gene Ther* **7**: 249–260.
- Yates, F and Daley, GQ (2006). Progress and prospects: gene transfer into embryonic stem cells. *Gene Ther* **13**: 1431–1439.
- Russell, DW and Hirata, RK (1998). Human gene targeting by viral vectors. *Nat Genet* **18**: 325–330.
- Gropp, M and Reubinoff, B (2006). Lentiviral vector-mediated gene delivery into human embryonic stem cells. *Meth Enzymol* **420**: 64–81.
- Fan, X, Valdimarsdottir, G, Larsson, J, Brun, A, Magnusson, M, Jacobsen, SE *et al.* (2002). Transient disruption of autocrine TGF-beta signaling leads to enhanced survival and proliferation potential in single primitive human hemopoietic progenitor cells. *J Immunol* **168**: 755–762.
- Hester, ME, Song, S, Miranda, CJ, Eagle, A, Schwartz, PH and Kaspar, BK (2009). Two factor reprogramming of human neural stem cells into pluripotency. *PLoS ONE* **4**: e7044.
- Jo, J and Tabata, Y (2008). Non-viral gene transfection technologies for genetic engineering of stem cells. *Eur J Pharm Biopharm* **68**: 90–104.
- Mansour, SL, Thomas, KR and Capecchi, MR (1988). Disruption of the proto-oncogene int-2 in mouse embryo-derived stem cells: a general strategy for targeting mutations to non-selectable genes. *Nature* **336**: 348–352.
- Kobayashi, K, Ohye, T, Pastan, I and Nagatsu, T (1996). A novel strategy for the negative selection in mouse embryonic stem cells operated with immunotoxin-mediated cell targeting. *Nucleic Acids Res* **24**: 3653–3655.
- Cathomen, T and Joung, JK (2008). Zinc-finger nucleases: the next generation emerges. *Mol Ther* **16**: 1200–1207.
- Lombardo, A, Genovese, P, Beausejour, CM, Colleoni, S, Lee, YL, Kim, KA *et al.* (2007). Gene editing in human stem cells using zinc finger nucleases and integrase-defective lentiviral vector delivery. *Nat Biotechnol* **25**: 1298–1306.
- Hockemeyer, D, Soldner, F, Beard, C, Gao, Q, Mitalipova, M, DeKaveler, RC *et al.* (2009). Efficient targeting of expressed and silent genes in human ESCs and iPSCs using zinc-finger nucleases. *Nat Biotechnol* **27**: 851–857.
- Zou, J, Maeder, ML, Mali, P, Pruett-Miller, SM, Thibodeau-Beganny, S, Chou, BK *et al.* (2009). Gene targeting of a disease-related gene in human induced pluripotent stem and embryonic stem cells. *Cell Stem Cell* **5**: 97–110.
- Ruby, KM and Zheng, B (2009). Gene targeting in a HUES line of human embryonic stem cells via electroporation. *Stem Cells* **27**: 1496–1506.
- Porteus, MH and Carroll, D (2005). Gene targeting using zinc finger nucleases. *Nat Biotechnol* **23**: 967–973.
- Miller, JC, Tan, S, Qiao, G, Barlow, KA, Wang, J, Xia, DF *et al.* (2011). A TALE nuclease architecture for efficient genome editing. *Nat Biotechnol* **29**: 143–148.
- Schaffer, DV, Koerber, JT and Lim, KI (2008). Molecular engineering of viral gene delivery vehicles. *Annu Rev Biomed Eng* **10**: 169–194.
- Maguire, AM, High, KA, Auricchio, A, Wright, JF, Pierce, EA, Testa, F *et al.* (2009). Age-dependent effects of RPE65 gene therapy for Leber's congenital amaurosis: a phase 1 dose-escalation trial. *Lancet* **374**: 1597–1605.
- Cideciyan, AV, Hauswirth, WW, Aleman, TS, Kaushal, S, Schwartz, SB, Boye, SL *et al.* (2009). Human RPE65 gene therapy for Leber congenital amaurosis: persistence of early visual improvements and safety at 1 year. *Hum Gene Ther* **20**: 999–1004.
- Vasileva, A, Linden, RM and Jessberger, R (2006). Homologous recombination is required for AAV-mediated gene targeting. *Nucleic Acids Res* **34**: 3345–3360.
- Jang, JH, Koerber, JT, Kim, JS, Asuri, P, Vazin, T, Bartel, M *et al.* (2011). An evolved adeno-associated viral variant enhances gene delivery and gene targeting in neural stem cells. *Mol Ther* **19**: 667–675.
- Smith-Arica, JR, Thomson, AJ, Ansell, R, Chiorini, J, Davidson, B and McWhir, J (2003). Infection efficiency of human and mouse embryonic stem cells using adenoviral and adeno-associated viral vectors. *Cloning Stem Cells* **5**: 51–62.
- Maheshri, N, Koerber, JT, Kaspar, BK and Schaffer, DV (2006). Directed evolution of adeno-associated virus yields enhanced gene delivery vectors. *Nat Biotechnol* **24**: 198–204.
- Koerber, JT, Jang, JH and Schaffer, DV (2008). DNA shuffling of adeno-associated virus yields functionally diverse viral progeny. *Mol Ther* **16**: 1703–1709.
- Koerber, JT, Klimczak, R, Jang, JH, Dalkara, D, Flannery, JG and Schaffer, DV (2009). Molecular evolution of adeno-associated virus for enhanced glial gene delivery. *Mol Ther* **17**: 2088–2095.
- Klimczak, RR, Koerber, JT, Dalkara, D, Flannery, JG and Schaffer, DV (2009). A novel adeno-associated viral variant for efficient and selective intravitreal transduction of rat Müller cells. *PLoS ONE* **4**: e7467.
- Excoffon, KJ, Koerber, JT, Dickey, DD, Murtha, M, Keshavjee, S, Kaspar, BK *et al.* (2009). Directed evolution of adeno-associated virus to an infectious respiratory virus. *Proc Natl Acad Sci USA* **106**: 3865–3870.
- Li, W, Zhang, L, Johnson, JS, Zhijian, W, Grieger, JC, Ping-jie, X *et al.* (2009). Generation of novel AAV variants by directed evolution for improved CFTR delivery to human ciliated airway epithelium. *Mol Ther* **17**: 2067–2077.
- Grimm, D, Lee, JS, Wang, L, Desai, T, Akache, B, Storm, TA *et al.* (2008). *In vitro* and *in vivo* gene therapy vector evolution via multispecies interbreeding and retargeting of adeno-associated viruses. *J Virol* **82**: 5887–5911.
- Wu, Z, Asokan, A and Samulski, RJ (2006). Adeno-associated virus serotypes: vector toolkit for human gene therapy. *Mol Ther* **14**: 316–327.
- Mitsui, K, Suzuki, K, Aizawa, E, Kawase, E, Suemori, H, Nakatsuiji, N *et al.* (2009). Gene targeting in human pluripotent stem cells with adeno-associated virus vectors. *Biochem Biophys Res Commun* **388**: 711–717.
- Khan, IF, Hirata, RK, Wang, P-R, Li, Y, Kho, J, Nelson, A *et al.* (2010). Engineering of human pluripotent stem cells by AAV-mediated gene targeting. *Mol Ther* **18**: 1192–1199.
- Müller, OJ, Kaul, F, Weltzman, MD, Pasqualini, R, Arap, W, Kleinschmidt, JA *et al.* (2003). Random peptide libraries displayed on adeno-associated virus to select for targeted gene therapy vectors. *Nat Biotechnol* **21**: 1040–1046.
- Park, IH, Zhao, R, West, JA, Yabuuchi, A, Huo, H, Ince, TA *et al.* (2008). Reprogramming of human somatic cells to pluripotency with defined factors. *Nature* **451**: 141–146.
- Kern, A, Schmidt, K, Leder, C, Müller, OJ, Wobus, CE, Bettinger, K *et al.* (2003). Identification of a heparin-binding motif on adeno-associated virus type 2 capsids. *J Virol* **77**: 11072–11081.
- Summerford, C and Samulski, RJ (1998). Membrane-associated heparan sulfate proteoglycan is a receptor for adeno-associated virus type 2 virions. *J Virol* **72**: 1438–1445.
- Miller, DG, Petek, LM and Russell, DW (2003). Human gene targeting by adeno-associated virus vectors is enhanced by DNA double-strand breaks. *Mol Cell Biol* **23**: 3550–3557.
- Pruett-Miller, SM, Connelly, JP, Maeder, ML, Joung, JK and Porteus, MH (2008). Comparison of zinc finger nucleases for use in gene targeting in mammalian cells. *Mol Ther* **16**: 707–717.
- Porteus, MH, Cathomen, T, Weltzman, MD and Baltimore, D (2003). Efficient gene targeting mediated by adeno-associated virus and DNA double-strand breaks. *Mol Cell Biol* **23**: 3558–3565.
- Khan, IF, Hirata, RK and Russell, DW (2011). AAV-mediated gene targeting methods for human cells. *Nat Protoc* **6**: 482–501.

41. Opie, SR, Warrington, KH Jr, Agbandje-McKenna, M, Zlotukhin, S and Muzyczka, N (2003). Identification of amino acid residues in the capsid proteins of adeno-associated virus type 2 that contribute to heparan sulfate proteoglycan binding. *J Virol* **77**: 6995–7006.
42. Levy, HC, Bowman, VD, Govindasamy, L, McKenna, R, Nash, K, Warrington, K et al. (2009). Heparin binding induces conformation changes in adeno-associated virus serotype 2. *J Struct Biol* **165**: 146–156.
43. Marcaida, MJ, Muñoz, IG, Blanco, FJ, Prieto, J and Montoya, G (2010). Homing endonucleases: from basics to therapeutic applications. *Cell Mol Life Sci* **67**: 727–748.
44. Johnson, RS, Sheng, M, Greenberg, ME, Kolodner, RD, Papaioannou, VE and Spiegelman, BM (1989). Targeting of nonexpressed genes in embryonic stem cells via homologous recombination. *Science* **245**: 1234–1236.
45. Schröder, AR, Shinn, P, Chen, H, Berry, C, Ecker, JR and Bushman, F (2002). HIV-1 integration in the human genome favors active genes and local hotspots. *Cell* **110**: 521–529.
46. Russell, DW and Hirata, RK (2008). Human gene targeting favors insertions over deletions. *Hum Gene Ther* **19**: 907–914.
47. Hasty, P, Crist, M, Grompe, M and Bradley, A (1994). Efficiency of insertion versus replacement vector targeting varies at different chromosomal loci. *Mol Cell Biol* **14**: 8385–8390.
48. Gupta, A, Meng, X, Zhu, LJ, Lawson, ND and Wolfe, SA (2011). Zinc finger protein-dependent and -independent contributions to the *in vivo* off-target activity of zinc finger nucleases. *Nucleic Acids Res* **39**: 381–392.
49. Petek, LM, Russell, DW and Miller, DG (2010). Frequent endonuclease cleavage at off-target locations *in vivo*. *Mol Ther* **18**: 983–986.
50. Yu, JH and Schaffer, DV (2006). Selection of novel vesicular stomatitis virus glycoprotein variants from a peptide insertion library for enhanced purification of retroviral and lentiviral vectors. *J Virol* **80**: 3285–3292.
51. Xie, Q, Bu, W, Bhatia, S, Hare, J, Somasundaram, T, Azzi, A et al. (2002). The atomic structure of adeno-associated virus (AAV-2), a vector for human gene therapy. *Proc Natl Acad Sci USA* **99**: 10405–10410.

	Espacio reservado para el adhesivo de radicación
---	--

DIRECCIÓN DE NUEVAS CREACIONES
Procedimiento acelerado de exámenes de patente (PPH)

1	NÚMERO DE SOLICITUD ANTE LA SIC	NC2018/0013255
2	OFICINA DE PRIMERA PRESENTACIÓN (OPP)	ESTADOS UNIDOS DE AMERICA
3	NÚMERO(S) DE SOLICITUD O PUBLICACIÓN DE LA(S) SOLICITUD(ES) (OPP):	1. WO/2017/197355 2. 3.
4	☐ REPRESENTANTE LEGAL ☒ APODERADO	
APELLIDOS CASTELLANOS ABONDANO		NOMBRES MARGARITA
		IDENTIFICACIÓN C.C . 39'779.654 T.P . 70819
DIRECCIÓN	CALLE 94A No. 7A-09	No. TELÉFONO 7560770
CIUDAD	BOGOTÁ	CORREO ELECTRÓNICO sic@castellanosyco.co
PAÍS	COLOMBIA	No. RADICACIÓN O PROTOCOLO DE PODER GENERAL
5	CONDICIONES PARA ACCEDER A ESTE PROCESO:	
5.1. DATOS DE PRIORIDAD ☒ SI ☐ NO		
(33) PAÍS DE ORIGEN 1. ESTADOS UNIDOS DE AMERICA 2. ESTADOS UNIDOS DE AMERICA 3. ESTADOS UNIDOS DE AMERICA 3. ESTADOS UNIDOS DE AMERICA		CÓDIGO PAÍS US US US US
		(31) NÚMERO 62/384,590 62/378,106 62/336,441 62/454,612
		(32) FECHA (AAAA/MM/DD) 2016/09/207 2016/08/22 2016/05/13 2017/02/03
5.2. DATOS SOLICITUD INTERNACIONAL PCT ☒ SI ☐ NO		
Solicitud Internacional No. PCT/US2017/032542		Fecha 12/05/2017
Publicación Internacional No. WO/2017/197355		Fecha 16/11/2017
El solicitante presenta una copia de la(s) reivindicación(es) patentable(s) u otorgable(s) de la solicitud de la Oficina Australiana de Patentes con el Nro. AU2017261812B2 que profirió el examen de patentabilidad, junto con una traducción al castellano.		
☒ SI ☐ NO		
REQUISITOS PARA EL RECONOCIMIENTO DE LOS RESULTADOS DE EXÁMENES DE PATENTABILIDAD EFECTUADOS A SOLICITUDES EXTRANJERAS		
Para que la Superintendencia de Industria y Comercio pueda reconocer los resultados de exámenes de patentabilidad producidos por otras Oficinas de Propiedad Industrial se debe cumplir con las siguientes condiciones:		
(a) Al momento de solicitar el examen de patentabilidad, si el solicitante pide que se reconozcan exámenes de patentabilidad de otras oficinas extranjeras, deberá aportar junto con el pago de la tasa correspondiente, los siguientes anexos:		
(i) Copia de las resoluciones, fallos o decisiones que han sido relevantes para la determinación de la patentabilidad de la(s) solicitud(es) en la Oficina que haya proferido el examen de patentabilidad que contienen reivindicaciones patentables u otorgables que son la base de la petición, y si es del caso, su traducción al castellano. ☒ SI ☐ NO		
(ii) Copia de todas las reivindicaciones determinadas como patentables u otorgables en la Oficina que profirió el examen de patentabilidad y, si es del caso, su traducción al castellano. ☒ SI ☐ NO		
(iii) Copia de los antecedentes, información o documentos citados por el examinador de la Oficina de Primera Presentación del examen de patentabilidad, en las resoluciones, fallos o decisiones correspondientes, incluyendo la literatura no patente. Los documentos de patente deberán presentarse siempre y cuando no estén disponibles para la Superintendencia de Industria y Comercio mediante el simple acceso a bases de datos gratuitas. ☒ SI ☐ NO		
(iv) Una tabla de correspondencias de las reivindicaciones que han sido determinadas como patentables u otorgables en la solicitud de la Oficina que haya proferido el examen de patentabilidad. ☒ SI		

	
Margarita Castellanos (Apoderada)	
7	FIRMA DEL SOLICITANTE, DEL APODERADO O DEL REPRESENTANTE LEGAL <i>Junto a cada firma, indicar el nombre del firmante y su calidad (si tal calidad no es obvia al leer el petitorio)</i>

6	TABLA DE CORRESPONDENCIA DE REIVINDICACIONES	
AU2017261812B2	Colombia	Correspondencia
1	1	Idéntica
2	2	Idéntica
3	3	Idéntica
4	4	Idéntica
5	5	Idéntica
6	6	Idéntica
7	7	Idéntica
8	8	Idéntica
9	9	Idéntica
10	10	Idéntica
11	11	Idéntica
12	12	Idéntica
13	13	Idéntica
14	14	Idéntica
15	15	Idéntica
16	16	Idéntica
17	17	Idéntica
18	18	Idéntica
19	19	Idéntica
20	20	Idéntica
21	21	Idéntica
22	22	Idéntica
23	23	Idéntica
24	24	Limitada a método <i>in vitro</i>
25	25	Limitada a método <i>in vitro</i>
26	-	Cancelada
27	-	Cancelada
28	-	Cancelada
29	-	Cancelada
30	-	Cancelada
31	-	Cancelada
32	-	Cancelada
33	-	Cancelada
34	-	Cancelada
35	-	Cancelada
36	-	Cancelada

Research article

Open Access

A genome survey of *Moniliophthora perniciosa* gives new insights into Witches' Broom Disease of cacao

Jorge MC Mondego^{†1}, Marcelo F Carazzolle^{†1}, Gustavo GL Costa¹, Eduardo F Formighieri¹, Lucas P Parizzi¹, Johana Rincones¹, Carolina Cotomacci¹, Dirce M Carraro², Anderson F Cunha³, Helaine Carrer⁴, Ramon O Vidal¹, Raíssa C Estrela¹, Odalys García¹, Daniela PT Thomazella¹, Bruno V de Oliveira¹, Acássia BL Pires⁵, Maria Carolina S Rio¹, Marcos Renato R Araújo¹, Marcos H de Moraes¹, Luis AB Castro⁶, Karina P Gramacho⁷, Marilda S Gonçalves⁸, José P Moura Neto⁸, Aristóteles Góes Neto⁹, Luciana V Barbosa¹⁰, Mark J Gultinan¹¹, Bryan A Bailey¹², Lyndel W Meinhardt^{†12}, Julio CM Cascardo^{†5} and Gonçalo AG Pereira^{*1}

Address: ¹Laboratório de Genômica e Expressão, Departamento de Genética e Evolução, Instituto de Biologia, Universidade Estadual de Campinas, CP 6109, 13083-970, Campinas – SP, Brazil, ²Laboratório de Genômica e Biologia Molecular, Hospital A.C. Camargo, 01509-010, São Paulo – SP, Brazil, ³HEMOCENTRO, Laboratório de Genoma e Hemoglobina, Universidade Estadual de Campinas, 13084-878, Campinas – SP, Brazil, ⁴Departamento de Ciências Biológicas, ESALQ, USP, 13418-900, Piracicaba – SP, Brazil, ⁵Laboratório de Genômica e Expressão Gênica, Departamento de Genética e Biologia Molecular, Universidade Estadual de Santa Cruz, 45650-000, Ilhéus – BA, Brazil, ⁶Embrapa Recursos Genéticos e Biotecnologia, Parque Estação Biológica – PqEB – Av. W5 Norte, 70770-900, Brasília – DF, Brazil, ⁷CEPLAC/CEPEC/SEFT, 45600-970, Itabuna – BA, Brazil, ⁸Laboratório de Biologia Molecular – Faculdade de Farmácia, Universidade Federal da Bahia, 40170-290, Salvador – BA, Brazil, ⁹Laboratório de Pesquisa em Microbiologia (LAPEM), Departamento de Ciências Biológicas, Universidade Estadual de Feira de Santana (UEFS), 44031-460, Feira de Santana – BA, Brazil, ¹⁰Laboratório de Biologia Molecular – Departamento de Biologia Geral, Instituto de Biologia, Universidade Federal da Bahia, 40170-290, Salvador – BA, Brazil, ¹¹Department of Horticulture, Pennsylvania State University, University Park, Chester, PA 16802, USA and ¹²Sustainable Perennial Crops Laboratory, USDA-ARS, 10300 Baltimore Av, Bldg. 001, 18 Beltsville MD 20705-2350, USA

Email: Jorge MC Mondego - jmcmondego@lge.ibi.unicamp.br; Marcelo F Carazzolle - mcarazzo@lge.ibi.unicamp.br; Gustavo GL Costa - glacerda@lge.ibi.unicamp.br; Eduardo F Formighieri - eduformi@gmail.com; Lucas P Parizzi - lucas@lge.ibi.unicamp.br; Johana Rincones - johanarp@unicamp.br; Carolina Cotomacci - caca@unicamp.br; Dirce M Carraro - dirce.carraro@hccancer.org.br; Anderson F Cunha - andfer@unicamp.br; Helaine Carrer - hecarrer@esalq.usp.br; Ramon O Vidal - vidal@lge.ibi.unicamp.br; Raíssa C Estrela - raissa@lge.ibi.unicamp.br; Odalys García - odalys@lge.ibi.unicamp.br; Daniela PT Thomazella - daniut@lge.ibi.unicamp.br; Bruno V de Oliveira - bvo@lge.ibi.unicamp.br; Acássia BL Pires - acassia@lge.ibi.unicamp.br; Maria Carolina S Rio - scatolin@lge.ibi.unicamp.br; Marcos Renato R Araújo - marcos.araujo@grialebiometrics.com; Marcos H de Moraes - mhmoraes@lge.ibi.unicamp.br; Luis AB Castro - lbarreto@mct.gov.br; Karina P Gramacho - karina@cepec.gov.br; Marilda S Gonçalves - mari@bahia.fiocruz.br; José P Moura Neto - jpneto@cpqgm.fiocruz.br; Aristóteles Góes Neto - agoesnt@uefs.br; Luciana V Barbosa - veiga@ufba.br; Mark J Gultinan - mjg9@psu.edu; Bryan A Bailey - bryan.bailey@ars.usda.gov; Lyndel W Meinhardt - lyndel.meinhardt@ars.usda.gov; Julio CM Cascardo - cascardo@uesc.br; Gonçalo AG Pereira^{*} - goncalo@unicamp.br

^{*} Corresponding author [†]Equal contributors

Published: 18 November 2008

Received: 17 June 2008

BMC Genomics 2008, 9:548 doi:10.1186/1471-2164-9-548

Accepted: 18 November 2008

This article is available from: <http://www.biomedcentral.com/1471-2164/9/548>

© 2008 Mondego et al; licensee BioMed Central Ltd.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/2.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Abstract

Background: The basidiomycete fungus *Moniliophthora perniciosa* is the causal agent of Witches' Broom Disease (WBD) in cacao (*Theobroma cacao*). It is a hemibiotrophic pathogen that colonizes

the apoplast of cacao's meristematic tissues as a biotrophic pathogen, switching to a saprotrophic lifestyle during later stages of infection. *M. pernicioso*, together with the related species *M. roreri*, are pathogens of aerial parts of the plant, an uncommon characteristic in the order Agaricales. A genome survey (1.9× coverage) of *M. pernicioso* was analyzed to evaluate the overall gene content of this phytopathogen.

Results: Genes encoding proteins involved in retrotransposition, reactive oxygen species (ROS) resistance, drug efflux transport and cell wall degradation were identified. The great number of genes encoding cytochrome P450 monooxygenases (1.15% of gene models) indicates that *M. pernicioso* has a great potential for detoxification, production of toxins and hormones; which may confer a high adaptive ability to the fungus. We have also discovered new genes encoding putative secreted polypeptides rich in cysteine, as well as genes related to methylotrophy and plant hormone biosynthesis (gibberellin and auxin). Analysis of gene families indicated that *M. pernicioso* have similar amounts of carboxylesterases and repertoires of plant cell wall degrading enzymes as other hemibiotrophic fungi. In addition, an approach for normalization of gene family data using incomplete genome data was developed and applied in *M. pernicioso* genome survey.

Conclusion: This genome survey gives an overview of the *M. pernicioso* genome, and reveals that a significant portion is involved in stress adaptation and plant necrosis, two necessary characteristics for a hemibiotrophic fungus to fulfill its infection cycle. Our analysis provides new evidence revealing potential adaptive traits that may play major roles in the mechanisms of pathogenicity in the *M. pernicioso*/cacao pathosystem.

Background

Moniliophthora pernicioso, previously known as *Crinipellis pernicioso* (Singer) Stahel, is a hemibiotrophic basidiomycete (Tricholomataceae, Agaricales, Marasmiaceae) fungus that causes Witches' broom disease (WBD) in cacao (*Theobroma cacao* L.) [1-3]. WBD and frosty pod rot (FPR), caused by *Moniliophthora roreri*, are the most devastating diseases of cacao in the Americas [4]. Cacao production in southeastern Bahia, the main production area in Brazil, was severely affected by the introduction of WBD at the end of 1980's [5]. This disease damaged Bahian agribusiness, caused major social problems and has contributed to the degradation of the Atlantic Rainforest ("Mata Atlântica"). This is because cacao producing areas, typically, maintained old-growth native tree species as shade for the cacao plantations, which were converted to pasture [6].

The symptoms displayed by cacao plants during WBD parallel the hemibiotrophic development of *M. pernicioso* [7]. Briefly, the disease begins when fungal spores germinate and infect meristematic tissues, developing into monokaryotic biotrophic hyphae without clamp connections that slowly occupy the intercellular space. This stage of WBD is characterized by the emergence of hypertrophic and hyperplastic anomalous branches, and the formation of parthenocarpic fruits. Infected branches, known as "green brooms", grow without apical dominance, with a phototropic orientation and displaying epinastic leaves [7]. After two to three months the infected tissue turns necrotic (dry brooms) and the hyphae become saprotrophic with two nuclei per cell and clamp connections,

invading the inter and intracellular spaces of the infected tissue [6,8]. This fungal species exhibits primary homothallism as its reproductive strategy [9]; thus, the change from the monokaryotic to the dikaryotic mycelium occurs without the prerequisite of mating between compatible individuals. After alternate wet and dry periods, basidiospores produced by the saprotrophic hyphae release basidiospores that are spread by wind or rain, thus completing the *M. pernicioso* life cycle [7].

The Witches' broom Genome Project <http://www.lge.ibi.unicamp.br/yassoura> involving several Brazilian laboratories was initiated to increase the knowledge of this disease. The genome size, chromosomal polymorphism, genetic variability and the *M. pernicioso* mitochondrial genome have already been described [10-12]. Additionally, a biochemical study revealed the metabolic modifications that occur in cacao plantlets during WBD development [13]. Technical improvements have been achieved in the manipulation of cacao [14,15] and *M. pernicioso* [16,17]. One of the main bottlenecks in *M. pernicioso* research was solved with the development of the *in vitro* production of biotrophic-like cultures [18]. Necrotic inducing proteins expressed by *M. pernicioso* have been characterized [19], and the analysis of EST libraries and DNA microarrays have identified differentially expressed genes during its development [20; A.B.L. Pires *et al.*, unpublished data] and for the interaction of the fungus with cacao [21,22]. Despite the substantial progression in understanding WBD, many questions remain unsolved, mainly those concerning the mechanisms controlling processes such as: (i) the fungal switch from biotrophism

to saprotrophism; (ii) the drastic phenotypic alteration of cacao during disease development and (iii) the death of infected tissues. *In vitro* cultures in our laboratory demonstrate that this fungus has a great capacity to adapt to media containing different sources of carbon and nitrogen and it is able to grow in extremely nutrient-poor media. These results together with the fact that hemibiotrophic fungi, such as *M. perniciosa*, display complex lifestyles [23] suggest that this fungus has a significant genomic and transcriptomic plasticity that contributes to the successful pathogenic mechanisms expressed during its life cycle. In a recent review, Meinhardt *et al.*, give a historical account and summarize the current state of knowledge about WBD [24].

Around two thirds of the known Basidiomycota species are included in the order Agaricales [25]. In addition to being an interesting group for carrying out developmental genetic studies concerning fungal development and reproduction, the Agaricales contain many important industrial species (i.e., edible mushrooms, fiber bleaching fungus), and species with unique lifestyles: saprophytes (i.e., wood-decaying fungus *Phanerochaete chrysosporium*), symbionts (i.e., ectomycorrhizal fungus *Laccaria bicolor*), leaf-litters decomposers (i.e., *Agaricus bisporus*) and root pathogens (i.e., *Armillaria mellea*). Interestingly, *M. perniciosa* and *M. roreri* are members of a group of Agaricales species that are able to infect aerial parts of plants, an uncommon characteristic among this Basidiomycota order.

Genome sequencing and analysis is an important strategy to obtain comprehensive information concerning the metabolism and development of organisms. The initial objective of the *M. perniciosa* Genome Project was to obtain a genome survey sequence using a whole shotgun strategy to provide genomic information for the WBD research community. Recently, the bioinformatics and genomic communities have been debating about the benefits and costs of finishing a complete genome as compared to applying a genome survey strategy [26-32]. It has been proposed that a two-fold genome sequence coverage is sufficient to support a high percentage of EST alignments and exon similarity matches [26,31]. Additionally, gene models resulting from a genome survey can be predicted accurately by the comparison with complete genomes of phylogenetically related organisms [26,31]. This strategy has been used for genomic surveys of dog (1.5× coverage) [31] and of the wine spoilage yeast *Dekkera bruxellensis* (0.4× coverage) [33]. Recently, the genomes of the basidiomycete species *Laccaria bicolor*, *Cryptococcus neoformans*, *Ustilago maydis*, *Coprinopsis cinerea* (*Coprinus cinereus*) and *Phanerochaete chrysosporium* have been determined and many additional Basidiomycota genome projects are ongoing or about to be released <http://www.broad.mit.edu/annotation/cgi>; <http://www.jgi.doe.gov>, which could provide a rich database for additional Basidiomycota genomic survey evaluations.

www.jgi.doe.gov, which could provide a rich database for additional Basidiomycota genomic survey evaluations.

With the existence of several complete Basidiomycota genomes, and both *M. perniciosa* ESTs libraries, and a 1.9 × genome coverage, we decided to conduct a genome survey of *M. perniciosa* to obtain further information about this important phytopathogen. This report describes a survey of the genome sequences of *M. perniciosa*, with specific emphasis on the genes potentially involved in disease development such as a cytochrome P450 monooxygenases, transposable elements, putative plant defense elicitors, pathogenicity effectors, cell wall degrading enzymes, proteins related to methylotrophy and the biosynthesis of plant hormones by the fungus. The identification of such genes in the *M. perniciosa* genomic data lead us to hypothesize a connection between molecular processes involved in the growth phases of the fungus and the progression of WBD.

Results and discussion

Genome assembly and estimation of genome size

A diagram flow describing the bioinformatic procedures applied in the *M. perniciosa* genome survey are depicted in Figure 1. After sequencing, a total of 124,565 reads were obtained and assembled using the whole genome shotgun strategy. During the initial assembly process, a large contig including 6,920 reads was found. This contig was identified as the mitochondrial genome of the fungus (RefSeq NC_005927), which comprised approximately 6% of *M. perniciosa* sequences [12]. The remaining genome sequences were assembled resulting in 17,991 contigs and 7,065 singlets with average lengths of 1,300 bp and 455 bp, respectively. The largest contig consisted of 25,364 bp, and was formed by 513 reads. The sum of all reads was close to 75 Mbp and the total assembly consisted of 26.7 Mbp (Fig. 1).

Previous Feulgen-image analysis experiments estimated *M. perniciosa* genome size to be 32.98 ± 7.95 Mbp [10]. Due to the large error in this estimation we decided to assess the genome size using the genome estimation protocol established in the dog genome survey [31] (more details in Additional File 1). This analysis resulted in a genome size ranging from 38.7 to 39.0 Mbp, a value similar to the genome length of another fungus belonging the order Agaricales, *C. cinerea* (37.5 Mbp), and that agrees with the previous size estimated by Feulgen-image analysis [10].

The Lander Waterman theory (LW) [34] and its applications [35,36] were used to confirm the estimate of the *M. perniciosa* genome size. The theoretical values for the expected number of clusters (contigs + singlets), contigs, gaps, average cluster length and average gap length can be calculated using the LW theory from the effective average

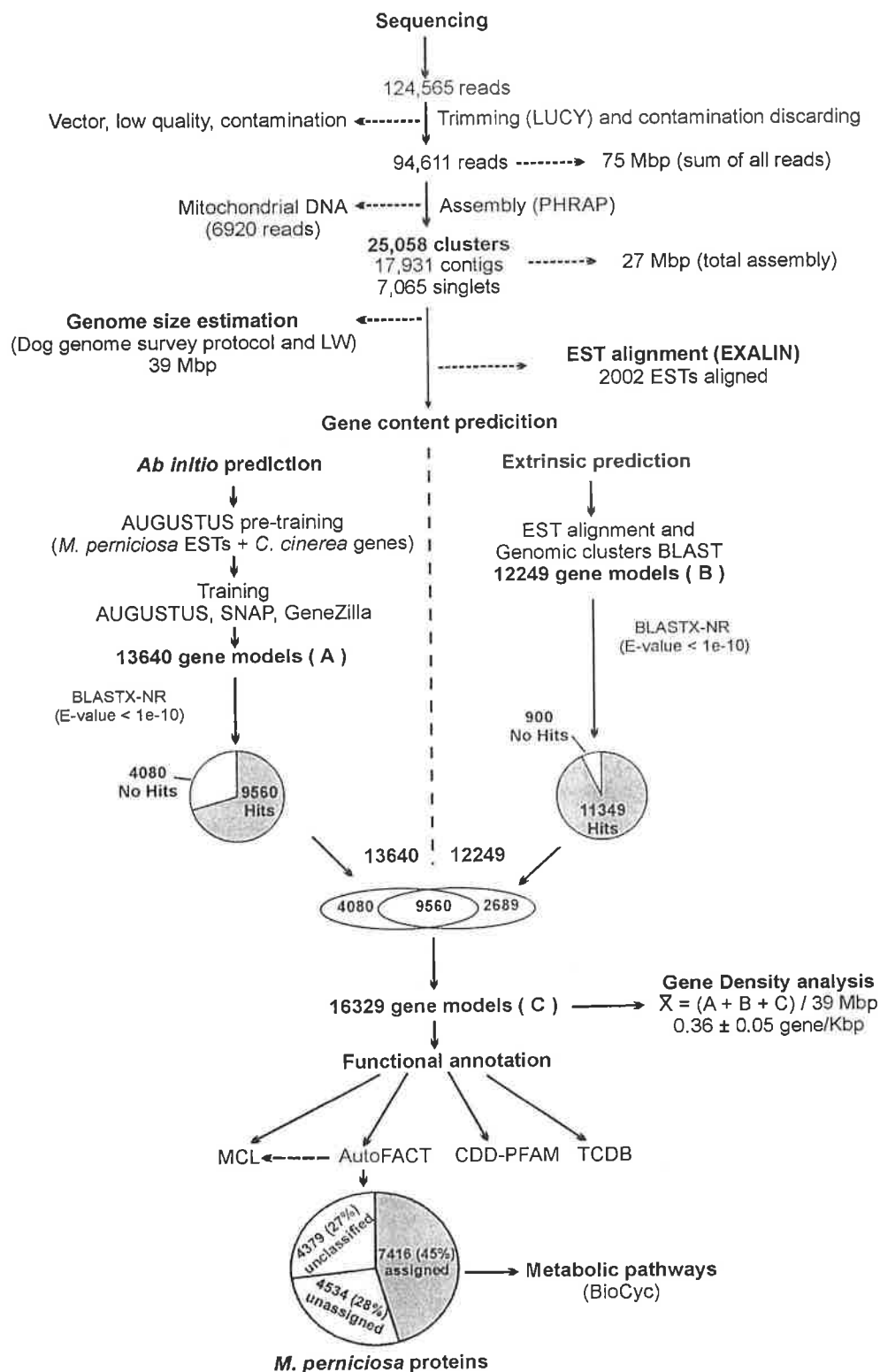


Figure 1
Flow diagram of bioinformatics procedures applied in *M. perniciosus* genome survey.

read length, the number of reads and the genome size as parameters. If the calculated values derived from the genome assembly were close to the theoretical values derived from LW calculations, then it is possible to confirm the estimated genome size. A summary of the results obtained from the assembly data and from the estimation using the LW theory is shown in Table 1. The experimental values agreed with the theoretical calculations, thus supporting our genome size estimation. According to LW analysis the average gap size was 413 pb. To have more information about the distribution of the gap size, we performed a comparison between a set of eukaryotic core proteins (generated by CEGMA pipeline [37]) and *M. perniciosus* contigs (See Additional file 1). Using this methodology, we detected that the average gap size was around 500 ± 300 bp, corroborating with LW average gap size estimate (413 bp). Considering that most of the gaps are around 500 bp and that *C. cinerea* average gene size is 1,678 bp, the majority of *M. perniciosus* genes or partial gene regions are likely to be included in the contigs. Corroborating this notion, we have found all genes encoding proteins of essential metabolic pathways such as Glycolysis, Gluconeogenesis, Pentose Phosphate pathway, and several others, which are available in the website <http://www.lgc.ibi.unicamp.br/biocyc/MP> (see below).

In order to estimate the number of sequences misassembled due to repeat regions, we applied the integrated pipeline for assembly validation, called amosvalidate [38] (Further information in Additional file 1). This analysis resulted in 664 contigs with overrepresented regions totaling 1.1 Mbp. Multiplying the number of bases in overrepresented regions by over-coverage estimate resulted in 7.4 Mbp of repeat regions in the genome. Since reads from repetitive regions were eliminated from the estimation of the genome length using dog genome survey protocol (Additional file 1), this misassembling does not invalidates the genome size estimate.

Gene content

The initial step to uncover the gene content of *M. perniciosus* was performed using the genomic sequences together with a library of 3,145 ESTs, previously annotated and partially published [20]. Genes were identified by comparing these two libraries using the program Exalin [39]. This analysis allowed us to identify expressed genes and intron structure. The result was 2,002 ESTs aligned to the

genome contigs. Based on this EST-genome sequence alignment, the average intron length was estimated as 52 bp. This information was then used in the next steps of the process.

A more detailed investigation of the gene content was carried out using a combination of *ab initio* (gene predictor programs) and comparative gene prediction (BLAST-EST sequence alignment). For *ab initio* gene prediction, we first applied the methodology described for the AUGUSTUS gene predictor [40] <http://augustus.gobics.de/binaries/retraining.html>. This methodology essentially trains the AUGUSTUS program using a combination of sequences from the species of interest together with sequences of a phylogenetically related species, specifically with similarities in intron and exon length distributions. Using this approach, the coding content sensors (codon usage, GC content) are trained with sequences of the species of interest, and signal sensors (splice sites, TATA-box, polyadenylation sites, etc) are trained with the sequences of the related species. Sequences from *C. cinerea* were used in the training process; together with selected *M. perniciosus* ESTs with E-value in BLASTX-NR $\leq 1E-10$. The sequences of the ESTs that aligned with proteins in the databank were concatenated, giving rise to a 240 Kbp sequence. Ten copies of this *M. perniciosus* EST concatamers (total of 2.4 Mbp) together with a dataset of genes from *C. cinerea* <http://augustus.gobics.de/datasets>, comprised of 1.2 Mbp, were used to "pre-train" AUGUSTUS. The *M. perniciosus* predictions that came out of this pre-training were compared with the protein databank NR using BLASTp. The resulting predictions with similarities in the NR databank, and with a coverage $\geq 90\%$, were selected. After eliminating redundancies, 134 complete (containing the start codon and stop codon) and 1,136 partial (without the start codon and/or stop codon) *M. perniciosus* gene models were used to train AUGUSTUS [40] and two other gene predictors (SNAP [41] and Genezilla [42]). Predictions with less than 30 amino acids were eliminated. The remaining predictions were grouped into 19,932 overlapping clusters; that is, genomic regions covered by at least one prediction. The predictions in each overlapping cluster were ranked according to the criteria used by the Fungal Genome Initiative at the Broad Institute <http://www.broad.mit.edu>, with some adaptations. The *ab initio* gene finding pipeline generated 13,640 gene models, 9,560 of which contained

Table 1: Comparison between assembly values and values calculated using Lander Waterman theory

	Calculated from assembling	Calculated from LW theory
# Clusters (contigs+singletons)	25,056	24,950
# Contigs	17,991	18,370
# Gaps	-	24,951
Average cluster size (bp)	1,065	1,152
Average gap size (bp)	-	413

significant similarity to GenBank sequences (Fig. 1 and Fig. 2).

The extrinsic prediction methodology consisted of a combination of genomic similarity searches (BLASTX) using contigs and singlets, with the alignment of *M. perniciosus* ESTs in the genomic clusters. 17,991 contigs and 7,065 singlets were submitted to similarity analysis in a databank containing BLASTX-NR plus *P. chrysosporium* proteins. The genomic regions containing homologues in this databank were selected and exon-intron boundaries were determined. We also selected the alignments between ESTs and genomic clusters. Then, we evaluated if there was a superposition of BLAST alignments and EST alignments. After this analysis, the extrinsic prediction methodology revealed 12,249 gene models. Most of these gene models (9,560) were also predicted by the *ab initio* gene predictor programs. Of the 2,689 remaining gene models not predicted by the gene predictor programs, 1,789 presented significant similarity to sequences deposited in the GenBank (E-value $\leq 1E-10$) and 900 did not have any significant similarity in the GenBank (E-value $> 1E-10$) (Fig. 1 and Fig. 2). One of the reasons why these genes were not detected by the gene prediction programs could be the presence of low quality sequences that may have lead to frameshifts, thus making them impossible to detect by these programs. The total number of gene models obtained using both approaches, *ab initio* and extrinsic predictions, is 16,329. Assuming that the average contig length is 1.3 Kbp and the average gap length is 413 bp, it is possible that this total number of gene models (16,329) represents an overestimation. For instance, in our assem-

bly a gene with > 2 Kbp (protein ~ 700 aa) could be represented by two gene models, with each one in different contigs, which suggests a possible redundancy in this gene model prediction. Thus, it is very likely that the real number of gene models will be less and closer to the *ab initio* or extrinsic predictions separately. However, to obtain the maximum amount of information, we decided to evaluate the results of similarity searches in GenBank (BLASTX-NR) for all 16,329 gene models. As a result, 69.5% (11,349) of the gene models showed significant sequence similarity (E-value $\leq 1e-10$) to GenBank proteins. The remaining 30.5% (4,980), designated as "no-hits", were sequences with lower E-value scores (E-value $> 1e-10$). Of these gene models 4,080 were predicted with the *ab initio* gene predictors and 900 with the extrinsic predictions (Fig. 1 and Fig. 2).

As expected, BLASTX analysis against available complete fungal genomes showed that *M. perniciosus* gene models present higher similarity to genes from Basidiomycota (*L. bicolor*, *C. cinerea* and *P. chrysosporium*) than those from Ascomycota (Table 2).

In order to find groups of similar proteins in our dataset, we applied a Markov Clustering (MCL) algorithm [43] to the *M. perniciosus* gene models. Although the output by this method is not highly reliable, they correlate well with "real" gene families and can be applied efficiently to cluster large quantities of genes in a high throughput fashion [44,45]. MCL also helped to the assemble genes without similarity in the GenBank into gene families, which are described in the following sections.

Those gene models that were not grouped into gene families by the MCL algorithm have been compared with the genome of *C. cinerea* in order to discard those genes with parts of their sequences present in more than one contig. In order to perform this comparison (TBLASTN) we assume that the length distribution of *M. perniciosus* proteins is similar to those from *C. cinerea*. In this comparison we evaluated similar genes according to the length of the protein (Fig. 3). This comparison showed that there is a clear correlation between complete genes in *C. cinerea* and *M. perniciosus* gene models coding for proteins smaller than 300 aa. However, for larger proteins more than one *M. perniciosus* gene model showed similarity to a single *C. cinerea* protein, thus confirming the overestimation of the number of gene models. Given the fact that the genome sequence is incomplete, there is high probability that we are predicting more than one gene model per gene.

In order to have a measurement of the overestimation and use it to correct our database, the corresponding relationship between the protein size and the number of gene models was calibrated with the genome of *C. cinerea* using

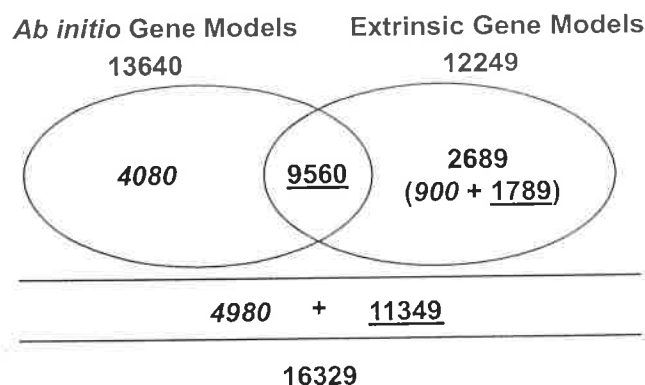


Figure 2
Number of *M. perniciosus* gene models predicted by *ab initio* and/or extrinsic prediction methods. Left ellipse: gene models predicted by *ab initio* methods. Right ellipse: gene models predicted by extrinsic methods. The intersection contains gene models detected by both methods. Underlined: number of gene models with BLASTX-NR E-value similarity $\leq 1e-10$. In italics: number of gene models with BLASTX-NR E-value similarity $> 1e-10$.

Table 2: *Moniliophthora perniciosa* predicted gene comparisons

Organism Comparison	Genes with* BLASTX hits	Genes with Top BLASTX Hits
<i>Laccaria bicolor</i>	9910 (60.6%)	4811 (29.27%)
<i>Coprinopsis cinerea</i>	10649 (65.1%)	3281 (19.97%)
<i>Phanerochaete chrysosporium</i>	9311 (57.0%)	2021 (12.30%)
<i>Fusarium graminearum</i>	5904 (36.1%)	57 (0.35%)
<i>Magnaporthe grisea</i>	5645 (34.5%)	39 (0.2%)
<i>Ustilago maydis</i>	5488 (33.6%)	42 (0.26%)
<i>Cryptococcus neoformans</i>	5454 (33.4%)	78 (0.47%)
<i>Neurospora crassa</i>	5127 (31.4%)	16 (0.10%)
<i>Saccharomyces cerevisiae</i>	3296 (20.2%)	4 (0.02%)
All NR database	11349 (69.5%)	—

* – Each *M. perniciosa* gene model was compared to all the proteins from the organisms listed in the table. BLASTX score was defined as $1e-10$

the gene models unlinked to gene families identified by the MCL algorithm (Fig. 3). A linear regression formula was used to correct for the overestimation according to the protein size. For example, proteins smaller than 300 amino acids will have only one model representing the gene, while larger proteins will have one model and a fraction of a second model representing its gene. This fraction represents the overestimation of the number of gene models.

In order to estimate the gene density in *M. perniciosa*, and deal with the overestimation evident by the previous analysis, we averaged the total number of gene models obtained by the different predictions: *ab initio* predictions (13,640), extrinsic predictions (12,249) and total number of gene predictions (16,329). This resulted in 14,072 gene predictions that were divided by the 39 Mbp estimated genome size, which gave a gene density of approximately 2.77 ± 0.37 Kbp/gene or 0.36 ± 0.05 gene/Kbp for the genome of *M. perniciosa*. Figure 4 depicts the gene density

comparison between *M. perniciosa*, two eubacteria (*Xylella fastidiosa* and *Escherichia coli*), one archae (*Haloquadratum walsbyi*), an apicomplexan (*Theileria parva*), a primitive chordate (*Ciona intestinalis*) and a series of fungal genomes, and this showed that our gene density evaluation is in agreement with gene density data of other filamentous fungi [46].

The overall *M. perniciosa* genomic features are summarized in Table 3. The genomic clusters of *M. perniciosa* were submitted to NCBI (GenomeProject ID 28951, Locus-tag prefix MPER) and the Whole Genome Shotgun project has been deposited at DDBJ/EMBL/GenBank under the project accession ABRE00000000. The version described in this paper is the first version, ABRE01000000.

Overall Functional Annotation and Metabolic maps

The number of gene models used for the functional annotation and building of the metabolic maps was 16,329, which represents the total number of gene models

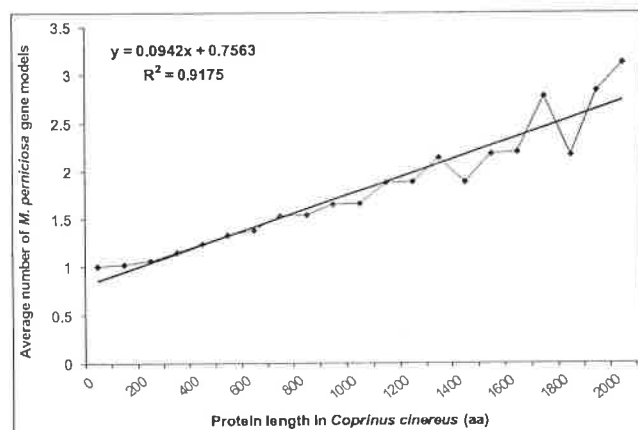


Figure 3
Correlation between the average number of *M. perniciosa* gene models and the length of *C. cinereus* proteins.

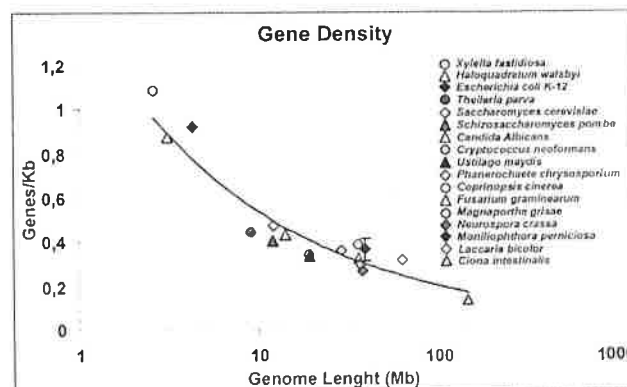


Figure 4
Correlation of various organisms' genome size and number of genes (Gene Density). Error bar in *M. perniciosa* data point depicts the Standard Deviation ($SD \pm 0.05$) of Gene/Kbp ratio using *ab initio* predictions, extrinsic predictions and the sum of both predictions (see text).

Table 3: *Moniliophthora perniciosa* genome survey features

Genome survey features	Values
Contigs	17,991
Contigs size (average)	1.3 Kbp
Singlets	7,065
Singlets size (average)	455 bp*
Sum of all reads	75 Mbp
Assembly size	26.7 Mbp
Estimated genome size	39 Mbp
Coverage	1.92×
Gene models	
<i>Ab initio</i> gene models	13,640
Extrinsic gene models	12,249
Total gene models	16,329
Gene density	2.77 Kbp/gene (SD ± 0.39)
	0.36 gene/Kbp (SD ± 0.05)
Average gene size	651.4 bp
GC content overall	47.7
GC content coding	49.7
GC content non coding	46.8
Intron size (mode)	52 bp
Intron size (average)	63 bp
Exon size (mode)	60 bp
Exon size (average)	166 bp
AutoFACT annotation	
Classified (Hits)	11,950 (73%)
Assigned (Conserved expressed proteins)	7,416 (45%)
Unassigned (Conserved hypothetical proteins)	4,534 (28%)
Unclassified (No Hits)	4,379 (27%)

* – Singlet size after LUCY quality trimming (phred ≥ 16).

obtained using *ab initio* and extrinsic predictions. Although this number represents an overestimation of the real number of genes (see above), it was used to maximize the information acquired from these sequences. All data obtained in the gene families and protein domain analyses were corrected according to the normalization procedure developed from the gene models not included in any of the MCL families based on comparisons with the *C. cinerea* genome (see above).

The program AutoFACT [47], an automated annotation tool, was used to evaluate the putative functions of *M. perniciosa* gene models. This program determines the most informative functional description by combining multiple BLAST reports from a number of user selected databases, and provides a consensus result [47]. AutoFACT classifies proteins as "classified proteins", which can be "assigned" (containing similarity to proteins with annotated function) or "unassigned" (containing similarity to proteins with unknown function), and "unclassified proteins" (without similarity to any other protein in data-banks). Using this software, 73% of gene models (11,950) were annotated as previously classified proteins. From these 7,416 (45%) were assigned proteins, and 4,534 (28%) were classified as unassigned proteins, which can

be interpreted as a class of conserved hypothetical proteins (Table 3). The remaining 4,379 gene models (27%) were considered as unclassified proteins. The number of gene models with E-value scores > 1e-10 obtained by BLASTX-NR (no-hits, see previous section) was 4,980 and represented 30.5% of the total number of models. The result obtained with AutoFACT increased the number of classified proteins by 3.5% (Fig. 1; Table 3). This indicates that the use of domain classification databases is helpful in the identification and annotation of gene models.

Additionally, we correlated the AutoFACT annotation with the gene families assembled by the MCL algorithm. As depicted in Table 4 and additional file 2, MCL data indicated that the cytochrome P450 monooxygenase family had the largest number of *M. perniciosa* members, followed by a gypsy-like retrotransposon gene family, and a gene family that has similarity to a *C. cinerea* hypothetical protein (EAU86912.1) and to a shitake mushroom *Len-tinula edodes* EST (EB016963). This new gene family appears to be related to specific developmental traits of the Agaricales. Also found were new gene families with unknown functions based on their lack of homology with GenBank sequences (Additional file 2; Additional File 3 – Worksheet Unknown Gene Families). These families

Table 4: MCL analysis of *M. perniciosus* gene models

MCL family ID	#members ^a	Norm Factor ^b	Norm #members ^c	Manual annotation ^d
1	175	1.24987	140	Cytochrome P450 monooxygenase
2	143	1.82632	78	Gypsy like-retrotranspon
4	84	1.35205	62	Conserved hypothetical protein
3	84	1.46889	57	Alcohol oxidases and GMC oxidoreductase
5	82	1.26479	65	Gypsy like-retrotranspon
6	76	1.57535	48	Serine threonine kinase
7	64	1.54312	41	Conserved hypothetical protein
8	59	1.54180	38	Conserved hypothetical protein
9	57	1.46370	39	Proteins containing WD-40 and NACHT domain
10	55	-	-	No hits protein family
11	52	1.34190	39	Cytochrome P450
12	50	1.52169	33	Serine threonine kinase
13	46	1.18098	39	Conserved hypothetical protein
14	43	1.21555	35	Cytochrome P450
15	41	1.45275	28	Carboxylesterase/lipase
16	39	1.67524	23	Gypsy like-retrotranspon
17	38	1.26598	30	Hexose transporter
19	35	1.11159	31	MFS transporter
18	35	1.43881	24	Splicing factor
20	32	1.17902	27	Laccase – Multicopper oxidase

a – Number of gene models present in each family

b – Normalization Factor used to normalize the number of gene models present in each family (see methods)

c – Normalization number of gene models present in each family

d – Manual annotation of MCL families according to BLAST similarity analysis

could possibly be linked to *M. perniciosus* physiological characteristics.

A comparison of the gene models with the CDD-PFAM databank [48] was performed to obtain information about protein domains present in *M. perniciosus* proteins. The data obtained were normalized using the procedure described above. Cytochrome P450 monooxygenase was the most prevalent protein domain assigned in *M. perniciosus* (gnl|CDD|40168), followed by protein kinases (gnl|CDD|40170), sugar transporters (gnl|CDD|40184), short chain dehydrogenases (gnl|CDD|40206) and carboxylesterases/lipases (gnl|CDD|40235) (Table 5; Additional File 4). The prevalence of Cytochrome P450 monooxygenase domains in the genome agrees with the results of MCL analysis. CDD-PFAM analysis was also used to evaluate the protein domains in other genome fungi (Additional File 4). The comparisons between *M. perniciosus* and the other fungi analyzed are described below.

A hypothetical metabolic map of *M. perniciosus* was built using BioCyc [49]. This analysis allowed us to annotate 235 metabolic pathways. These include 1358 enzymatic reactions incorporating a total of 2139 enzymes <http://www.lgc.ibi.unicamp.br/biocyc/MP>. A comparison of this metabolic map with the fungal model *Saccharomyces cere-*

visiae S288C (documented with 132 pathways, 925 enzymatic reactions, and 675 enzymes – <http://www.lgc.ibi.unicamp.br/biocyc/comparatives/MP-SCER>) showed that *M. perniciosus* has more metabolic pathways than *S. cerevisiae*; a result that corresponds to the smaller genome size of *S. cerevisiae* and possibly the more complex lifestyle of *M. perniciosus*. Interestingly, *M. perniciosus* has a higher number of reactions involving O₂, CO₂, H₂O₂, and NAD(P)+/NAD(P)H than *S. cerevisiae*, suggesting a greater capacity to use and deal with oxidation-reduction reactions (Additional File 5 – worksheet compounds). We also detected reactions with farnesyl pyrophosphate and dimethylallyl-diphosphate, which are involved in the biosynthesis of the secondary metabolites such as isoprenoids and indoles. Moreover, according to BioCyc analysis, *M. perniciosus* has more amino acid catabolic pathways, alternative carbon sources degradation and biosynthesis routes and C1 compounds (i.e., methanol) utilization and assimilation reactions than *S. cerevisiae* (Additional File 5 – worksheet pathways). These pathways are under manual annotation and will be published on BioCyc web page. The existence of these pathways in *M. perniciosus* suggest ecological and physiological adaptations to environmental stresses; to competition present in its native habitat in the Amazon Basin, and to traits that enable it to colonize cacao and trigger WBD.

Table 5: Top 20 CDD-PFAM domains in *M. perniciosus* proteins

CDD - ID ^a	Domain PFAM ^b	#Hits ^c Domains	%Hits ^d PTN	Rank ^e	Norm Factor	# Hits ^f Domains Norm	%Hits ^h PTN Norm	Rank ⁱ Norm
gnl CDD 40168	pfam00067, Cytochrome P450	256	1.57%	1	1.3565	188.71427	1.15%	1
gnl CDD 40170	pfam00069, Protein kinase	141	0.86%	2	1.6059	87.80184	0.54%	2
gnl CDD 40184	pfam00083, Sugar (and other) transporters	89	0.54%	3	1.6902	52.65651	0.32%	3
gnl CDD 40206	pfam00106, adh_short, short chain dehydrogenase	84	0.51%	4	1.7141	49.00442	0.30%	4
gnl CDD 40235	pfam00135, COesterase, Carboxylesterase	57	0.35%	5	1.4738	38.67603	0.24%	5
gnl CDD 40207	pfam00107, ADH_zinc_N, Zinc-binding dehydrogenase	43	0.26%	8	1.2398	34.68221	0.21%	6
gnl CDD 40813	pfam00732, GMC_oxred_N, GMC oxidoreductase	52	0.32%	6	1.5402	33.76192	0.21%	7
gnl CDD 40345	pfam00248, Aldo_ket_red, Aldo/keto reductase family	41	0.25%	9	1.2881	31.83088	0.19%	8
gnl CDD 41245	pfam01185, Hydrophobin, Fungal hydrophobin	26	0.16%	21	0.8460	30.73119	0.19%	9
gnl CDD 40253	pfam00153, Mito_carr, Mitochondrial carrier protein	38	0.23%	12	1.3890	27.35860	0.17%	10
gnl CDD 40748	pfam00665, rve, Integrase core domain	50	0.31%	7	1.8943	26.39449	0.16%	11
gnl CDD 40493	pfam00400, WD40, WD domain, G-beta repeat	35	0.21%	13	1.5151	23.10130	0.14%	12
gnl CDD 40179	pfam00078, RVT, Reverse transcriptase	39	0.24%	11	1.7620	22.13448	0.14%	13
gnl CDD 45101	pfam05199, GMC_oxred_C, GMC oxidoreductase	28	0.17%	16	1.2741	21.97654	0.13%	14
gnl CDD 40588	pfam00501, AMP-binding, AMP-binding enzyme	40	0.24%	10	1.8537	21.57819	0.13%	15
gnl CDD 41412	pfam01360, Monooxygenase, Monooxygenase	25	0.15%	22	1.2253	20.40382	0.12%	16
gnl CDD 40107	pfam00005, ABC_tran, ABC transporter	33	0.20%	14	1.7198	19.18851	0.12%	17
gnl CDD 40367	pfam00270, DEAD, DEAD/DEAH box helicase	30	0.18%	15	1.5947	18.81230	0.12%	18
gnl CDD 40419	pfam00324, AA_permease, Amino acid permease	26	0.16%	20	1.4641	17.75798	0.11%	19
gnl CDD 41783	pfam01753, zf-MYND, MYND finger	25	0.15%	23	1.4288	17.49737	0.11%	20

a - Number of CDD entry

b - Number and description of PFAM domain

c - Number of gene models containing each CDD-PFAM domain

d - Percentage of gene models containing each CDD-PFAM domain in relation to total number of *M. perniciosus* gene models (proteins).

e - Non-normalized CDD-PFAM ranking

f - Factor used to normalize the number of gene models containing each CDD-PFAM domain (see methods)

g - Normalized number of gene models containing a CDD-PFAM domain

h - Percentage of normalized number of gene models containing each CDD-PFAM domain in relation to total number of *M. perniciosus* gene models (proteins)

i - Normalized *M. perniciosus* CDD-PFAM ranking

Detoxification and general resistance mechanisms: cytochrome P450 monooxygenases, efflux transporters and anti-oxidative apparatus

Based on CDD-PFAM and MCL analyses gene members of the cytochrome P450 monooxygenase superfamily are prevalent in the genome of *M. perniciosa* (Tables 4 and 5; Additional File 3 – Worksheet P450; Additional File 4). Cytochrome P450 monooxygenases play a role in hydroxylation and oxidation processes involved in biosynthesis and degradation of different compounds [50]. Therefore, a large number of gene models similar to cytochrome P450 monooxygenases suggest a significant capacity for synthesis of secondary metabolites, such as hormones or toxins, and for detoxification. Among the fungi analyzed, *M. perniciosa* has the highest number of cytochrome P450 monooxygenase genes (188 gene models) representing 1.15% of the gene models (Fig. 5A). Saproprotrophic basidiomycetes *P. chrysosporium* and *C. cinerea* and the hemibiotrophic ascomycetes *Magnaporthe grisea* and *Fusarium graminearum* also have more than one hundred cytochrome P450 monooxygenase genes representing between 1.25% and 0.9% of their gene models (Figure 5A; Additional File 4). The basidiomycete *L. bicolor* (an ectomycorrhizal fungus) and the ascomycete *Neurospora crassa* (a fire-scoured landscape colonizer) follow with fewer genes. *U. maydis* (a biotrophic pathogen), *C. neoformans* (an animal pathogen) and *S. cerevisiae* (a fermentative fungus) have the fewest cytochrome P450 monooxygenase genes of the fungi compared in this study (Fig. 5A; Additional File 4). This analysis clearly demonstrates the prevalence of cytochrome P450 monooxygenases in saprotrophic and hemibiotrophic fungi, which have to hydrolize complex wood polymers and deal with a highly oxidative environment. As discussed by Gonzalez and Nebert [50], cytochrome P450 monooxygenase polymorphisms may be the product of the "molecular warfare" that occurs during the co-evolution of preys and predators, which produce toxins and detoxifying genes, respectively. This logic can be extrapolated to the plant-fungus interaction, and in that sense, we believe that the plethora of cytochrome P450 monooxygenases in *M. perniciosa* may be critical to detoxification and environmental adaptation as well as for disease development.

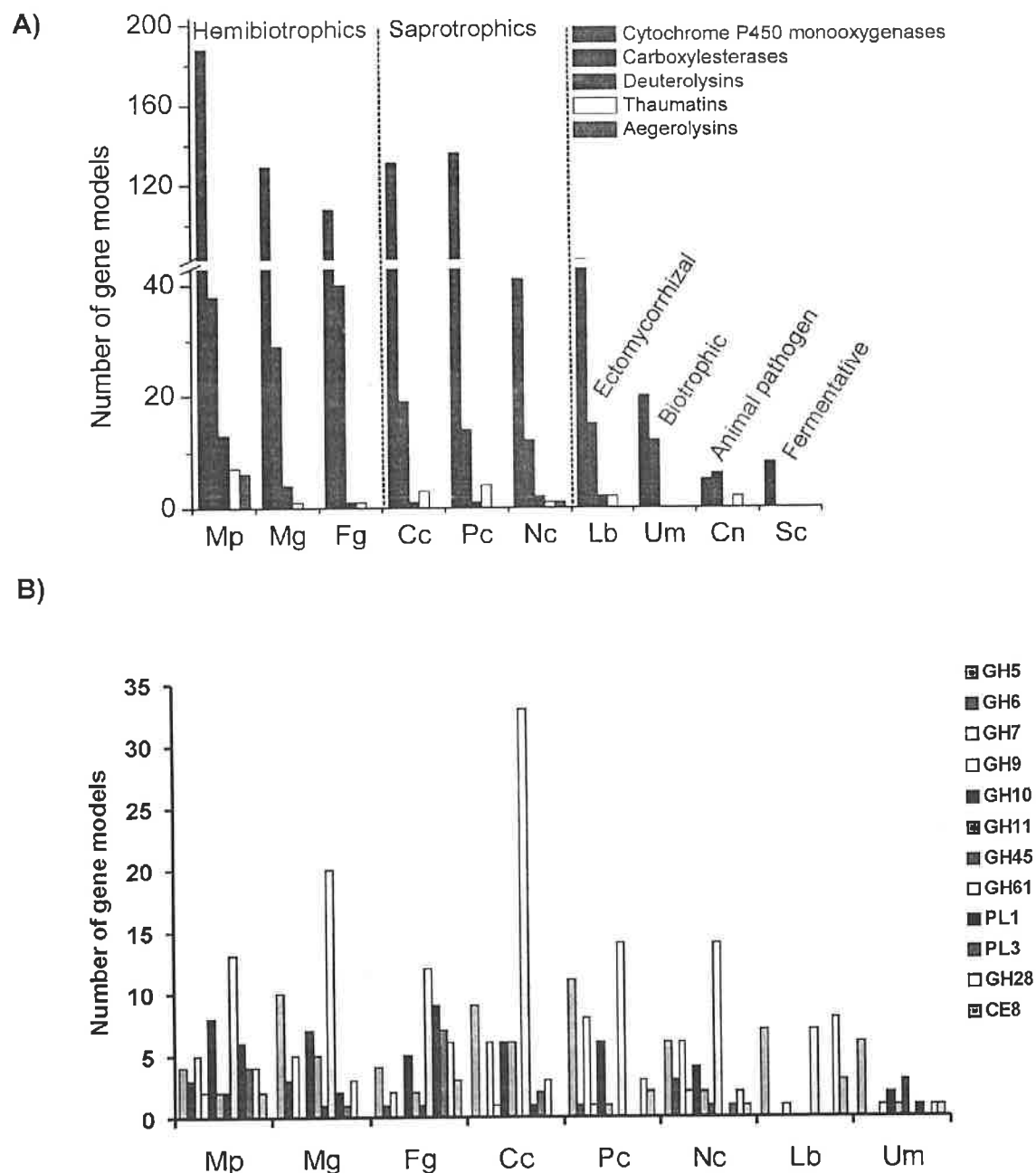
Another set of proteins related to detoxification processes are the efflux transporters. Similarity searches in the Transport Classification Database (TCDB – <http://www.tcdb.org>) [51] which compare the genome of *M. perniciosa* with other fungal genomes (Additional File 3 – Worksheet Functional annotation and Worksheet Transporters) verified an extended set of efflux transporters from classes 3.A.1 (ABC superfamily) and 3.A.3 (P-type ATPase Superfamily). The majority of ABC transporters identified in *M. perniciosa* belong to the Pleiotropic Drug Resistance (PDR) family and the ABC Conjugate (ABCC)

Transporter family. These proteins have been associated with fungal pathogenesis [52] and heavy metal resistance [53]. The members of P-type ATPase superfamily in *M. perniciosa* include phospholipid translocating ATPases family and fungal ENA-ATPases transporters, which are involved with the efflux of excessive Na⁺, and especially K⁺, encountered by fungi during colonization of plants [54].

During the plant defense, reactive oxygen species (ROS) are produced to limit the pathogen invasion [55]. However, pathogenic fungi produce antioxidant enzymes that enable them to neutralize host ROS. *M. perniciosa* contains a myriad of O₂⁻ and H₂O₂ decomposing enzymes such as superoxide dismutases, catalases, peroxiredoxins, glutathione-system, thioredoxin-system enzymes and manganese dependent peroxidases (Additional File 3 – Worksheet Functional Annotation). Therefore, this fungal genome harbors a complete ROS detoxification system. Reports indicate that H₂O₂ favors necrotrophic pathogenesis infection [56,57]. Furthermore, *M. perniciosa* produces calcium oxalate crystals (COC) [58], and a cacao susceptible genotype accumulates COC during *M. perniciosa* infection, followed by a programmed cell death (PCD) [59]. The degradation of COC produces carbon dioxide and H₂O₂, suggesting that COCs can be important to necrotrophic mycelia development. In addition, oxalate chelates Ca²⁺, an important secondary plant defense messenger and a key cross-linker of pectin in the middle lamella pectin [60], and was found to be a trigger of PCD in plants [61]. These findings suggest that oxalate favors *M. perniciosa* infection by disorganizing plant defense and plant cell wall structure, by facilitating the action of fungal pectinases and possibly triggering PCD in the later stages of WBD [59].

Genome variability: Mating-type genes and transposable elements

Seven *M. perniciosa* gene models were found to be similar to pheromone receptors (Additional File 3 – Worksheet Functional Annotation). As a primary homothallic fungus, *M. perniciosa* does not use its mating type system to outcross, but probably to promote the formation of clamp connections, hyphae dikaryotization and for the expression of pathogenicity genes as in *U. maydis* [62]. Previous reports have indicated that *M. perniciosa* exhibits high genetic variability at the molecular level [11,63,64]. This level of variability may be the reason *M. perniciosa* overcomes resistant genotypes of cacao, such as Scavina 6 [65]. Furthermore, it has been postulated that the genome variability found in homothallic *M. perniciosa* may be due to transposable elements (TEs) and ectopic recombination guided by the numerous copies of these elements found in the genome [10]. The fact that retrotransposons were identified in EST libraries and differentially expressed dur-

**Figure 5**

Comparison of *M. perniciosa* protein families with other fungi. (A) comparison between cytochrome P450 monooxygenases, carboxylesterases, deuterolysins, thaumatin and aegerolysins; (B) comparison of plant cell wall degrading enzymes from fungi that interact with plants. Mp = *Moniliophthora perniciosa*, Lb = *Laccaria bicolor*, Cc = *Coprinopsis cinerea*, Pc = *Phanerochaete chrysosporium*, Um = *Ustilago maydis*, Mg = *Magnaporthe grisea*, Cn = *Cryptococcus neoformans*, Fg = *Fusarium graminearum*, Nc = *Neurospora crassa*, Sc = *Saccharomyces cerevisiae*. The legend at the right refers to the nomenclature of plant cell wall degrading enzymes according to CAZy <http://www.cazy.org>.

ing *M. perniciosus* development [20], indicates that they are active elements, which could contribute to genetic variability. Among the *M. perniciosus* TE families, Gypsy-like retrotransposons were the most abundant, followed by Copia-like retrotransposons (Tables 4 and 5; Additional File 2; Additional File 3 – Worksheet Transposons). TEs similar to *P. chrysosporium* Copia-like elements were found in *M. perniciosus* genome inserted within putative cytochrome P450 monooxygenase genes [66]. Curiously, they were also inserted in a *P. chrysosporium* cytochrome P450 monooxygenase subfamily (Additional file 2 – family 255; Additional File 3 – Worksheet P450), suggesting that these TEs may have a common ancestral origin in Basidiomycota. Retroelements of the tyrosine recombinase (YR) order [67] and DNA transposons (class II transposons) from CACTA [68], hATC [69,70] and Tc/Mariner [71] superfamilies were also found. Finally, a previously described *Boto* DNA transposon [M.V. Queiroz *et al.*, unpublished data] from the PIF/IS5 superfamily [72] was also identified in a MCL family (Additional File 2 – family 251; Additional File 3 – Worksheet Transposons).

Plant Hormonal Disarrangement: Fungal genes related to plant hormones biosynthesis

There is growing evidence of phytohormones being produced by pathogens during some infective processes [73]. For instance, the production of gibberellins (GA); hormones involved in the regulation of stem elongation, seed germination, flowering and fruit maturation; have been identified in phytopathogenic bacteria and fungi that cause overgrowth symptoms, such as *Giberella fujikuroi* and *Sphaceloma manihoticola* [74,75]. A search for homologues of the fungal specific bi-functional ent-kaurene synthase (CPS/KS) responsible for the two-step cyclization of GGDP in fungi [76] identified gene models similar to the N-terminal domain of *G. fujikuroi* CPS/KS but did not detect any sequence similar to the C-terminal domain of this protein in *M. perniciosus*. Another gene model similar to CPS/KS that lacks the C-terminal domain was found in the *Aspergillus niger* genome (AM270241.1). Genes similar to GA-4 desaturase and GA oxidases (cytochrome P450), part of a GA biosynthesis gene cluster present in *G. fujikuroi* and *Phaeosphaeria* sp. were detected (Additional File 3 – Worksheet Functional Annotation). Reinforcing our data mining discover, is the fact that a gibberellin-like compound was detected in basidiospores of *M. perniciosus* [77]. We can theorize that the production of GA by *M. perniciosus* may confer the hyperplastic phenotype of the green broom that resembles stem hyper-elongation caused by GA-producing phytopathogens.

Another interesting discovery in the *M. perniciosus* genome is the presence of genes encoding enzymes of two biosynthetic pathways of indole-3-acetic acid (IAA), the most abundant natural plant auxin. We found a gene similar to

plant nitrilases (E.C 3.5.5.1) which catalyzes the direct conversion of indole-3-acetonitrile into ammonia and IAA. Additionally, potential genes for the IAA-producing tryptamine pathway (one tryptophan decarboxylase, copper amine oxidases and a hypothetical indole-acetaldehyde oxidase) were found (Fig. 6; Additional File 3 – Worksheet Functional Annotation). Furthermore, a manual annotation of EST libraries reveal the presence of an aromatic amino acid aminotransferase, which could make part of Indole-3-pyruvate IAA biosynthetic pathway (Fig. 6). IAA regulates many plant biological processes including cell elongation and fruit ripening. Recently, the presence of IAA in *M. perniciosus* basidiocarps was reported [78]. IAA is produced by other fungal phytopathogens such as biotrophics *U. maydis* and *Taphrina deformans* and hemibiotrophic *Colletotrichum gloeosporioides* sp. [79-81], and induces filamentation and invasive growth in *S. cerevisiae* [82]. Curiously, both Gibberellin and IAA induce fruit parthenocarp and act synergistically in plant organ expansion [83], both of which are traits of WBD.

Cacao plantlets inoculated with *M. perniciosus* spores emit higher levels of ethylene during the late stages of infection than non-inoculated plants [13]. The plant hormone ethylene stimulates elongation at low concentrations, and senescence, fruit ripening, and epinasty at high doses [84]. Auxins stimulate the synthesis of ethylene [85], which together with gibberellins have integrated actions in plant cell death [86] and stem elongation during phytochrome-mediated shade avoidance, a phenomenon that occurs in response to the low red to far-red light ratios (R:FR) under dense canopies [87]. We hypothesize that in a dense and shaded environment, such as a cacao plantation, the low R:FR ratio effects can be increased by the action of the aforementioned phytohormones, explaining the shade

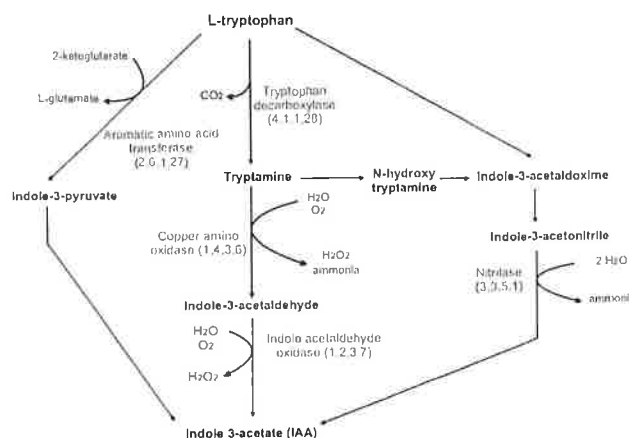


Figure 6
Indole-3-acetate (IAA) biosynthesis pathways. *M. perniciosus* gene models are annotated in red.

avoidance, hypertrophy, and elongation of green brooms. The presence of genes related to plant hormones production in *M. perniciosa* supports previous data [77,78] and suggests that this fungus can influence the plant metabolism and defense, by altering hormonal balance during infection [73,88].

Pathogenicity: Fungal effectors and pathogenicity associated proteins

As a pathogen that colonizes the plant apoplast during its biotrophic stage, *M. perniciosa* may release elicitor or effector proteins into the extracellular medium, which in turn could evade or suppress the plant defense response. Throughout WBD, *M. perniciosa* produces proteins with the potential to kill plant cells, thereby releasing their contents, which are absorbed by the fungus during its saprotrophic stage. Thus, *M. perniciosa* uses a varied arsenal of effector proteins in order to complete its infection cycle. Although some phytopathogenic fungi deliver effector proteins into the cytoplasm by means of haustoria [89,90] this type of structure is absent in *M. perniciosa*. Therefore it is possible that this fungus secretes these proteins into the apoplast, as has been described for other fungi [91]. Effector proteins that are recognized by plant resistance (R) proteins are known as *avr* proteins. No orthologues to the known Ascomycota *avr* genes, not even the *U. maydis* genes contained in "biotrophic clusters" [92], were found in the genome of *M. perniciosa*. However, *ab initio* gene prediction with peptide signal analysis revealed 70 "no hits" small proteins containing secretion signals and at least two cysteines (see methods, Additional File 6), which is a common trait of many proteins that are delivered into the host apoplast by phytopathogens [93]. In order to validate 22 selected *ab initio* predictions, we conducted RT-PCR using RNA from *M. perniciosa* saprotrophic mycelia. We validated the expression of 13 gene models out of 22 tested, all of which contained secretion signals (Additional File 7 – Primers No Hits Cys protein). Possibly, the gene models that were not confirmed by RT-PCR are expressed in other developmental stages of the fungus (basidiome, spore, etc.). Additional File 8 depicts the amplification of three of these genes. Whether these proteins play a role in the pathogenicity of *M. perniciosa* or in elicitation of cacao defense remains to be elucidated.

M. perniciosa contains Necrosis and Ethylene inducing proteins (NEPs) and cerato platanins [19,94], which can act in conjunction with a series of proteinases, hemolysin-like proteins and carboxylesterases/lipases found in the genome (Additional File 3 – Worksheet Functional Annotation). These proteins appear to be part of the destructive arsenal of *M. perniciosa*. The most abundant proteinases in the genome of *M. perniciosa* are deuterolysins, a type of fungal metalloproteinases that are similar to bacterial thermolysin [95]. Compared to other fungi indicates that *M. perniciosa* has a deuterolysin expansion (13 gene

models; Fig. 5A; Additional File 4), suggesting an important role for these proteinases during this fungus development.

We also identified a gene family similar to agaricales *Pleurotus ostreatus* and *Agrocybe aegerita* hemolysin-like aegerolysins. These proteins have cytolytic properties [96] and seems to play an important role at the initial phase of fungal fruiting by making the fungal membranes permeable during cell signaling [96]. *M. perniciosa* genome contain a family of aegerolysins (6 aegerolysins; Fig. 5A; Additional File 4). None of the other agaricales fungi analyzed (*L. bicolor*, *C. cinerea* and *P. chrysosporium*) contain these proteins. Even though we could not assess the genomic data of the basidiomycete containing aegerolysins, we suggest that the diversification of these proteins in WBD causative agent indicates their importance in *M. perniciosa* development or even in fungi defense and infective process.

Carboxylesterases and lipases are overrepresented in *M. perniciosa* (Tables 4 and 5; Additional Files 2 and 4). According to CDD-PFAM, *M. perniciosa* have 38 gene models annotated as carboxylesterases, approximately twice the number of such proteins in other basidiomycete (ectomycorrhizal *L. bicolor* (15); saprobes *C. cinerea* (19) and *P. chrysosporium* (14); and biotrophic *U. maydis* (12); Figure 5A; Additional File 3); while the hemibiotrophic ascomycete *F. graminearum* and *M. grisea* have 40 and 29 carboxylesterases, respectively. We postulate that the great number of carboxylesterases and lipases in *M. perniciosa*, *F. graminearum* and *M. grisea* is related to their hemibiotrophic lifestyle. In fact, these enzymes are induced during carbon and nitrogen starvation [97], and cell wall degradation [98], two events that occur during hemibiotrophism.

Other genes associated with the plant-pathogen interaction found in our analysis were similar to the SCP-like superfamily proteins, which comprise pathogenesis related (PR) proteins of family 1 (PR-1). Additionally, gene models similar to PR-5/thaumatin superfamily were also detected in the *M. perniciosa* genome (Additional File 3 – Worksheet Functional Annotation). PR proteins are well described as associated with defense reactions in plants against various pathogens [99]. For instance, transgenic plants overexpressing PR-1 proteins were more resistant to oomycete infection [100] and some Thaumatin-like proteins (TLPs) have β -glucanase activity, inhibit xylanase and have antifungal properties [101-103]. Recently, proteins similar to PR-1 and thaumatin have been characterized in animals and fungi [104,105], indicating a conserved and important role in diverse organisms. Based on CDD-PFAM analysis, *M. perniciosa* contains the largest number of thaumatins of any fungus sequenced, so far (7 thaumatins, Fig. 5A; Additional File 4).

Both PR-1 and PR-5 are induced by salicylic acid (SA) in plants [99]. Curiously, the *M. perniciosus* necrotrophic (saprotrophic) mycelia were found to produce and have tolerance to SA in axenic cultures [70]. Plants with WBD have a higher content of SA when compared with healthy plants [106]. *M. perniciosus* tolerance to SA could be explained by the expression of genes encoding salicylate hydroxylases, which were also detected in the genome (Additional File 3 – Worksheet Functional Annotation). In this scenario, high levels of SA could block the synthesis of jasmonic acid (JA), a defense compound against necrotrophic pathogens that acts as a necrosis inducer, thus rendering the plant susceptible to the spread of *M. perniciosus* [78,106]. In addition, it is possible that SA and SA-induced proteins (i.e., PR-1, PR-5) may act to limit competition from other microbial competitors during WBD progression, which would be an important component of the *M. perniciosus* pathogenicity strategy.

Colonization: Plant cell wall degrading enzymes (PCWDE)

Degradation of hemicellulose, cellulose, pectin and depolymerization of lignin are some of the mechanisms that necrotrophic fungi use to colonize plant tissues [107]. We identified genes encoding enzymes involved in degradation of hemicellulose and cellulose, including β -1,4 cellulases, exocellobiohydrolases, endo-beta-1,4-xylanases and endoglucanases; genes encoding lignolytic enzymes including manganese dependent peroxidase and multi-copper polyphenoloxidases (laccases); and genes encoding enzymes involved in pectin degradation, such as pectate lyases, polygalacturonases (pectinases) and pectin methylesterases (pectinesterase) (Additional File 3 – Worksheet Functional Annotation). *M. perniciosus* have an arsenal of plant cell wall degrading enzymes that is similar to that found in the hemibiotrophic pathogens *F. graminearum* and *M. grisea* (Fig. 5B; Additional File 9). Biotrophic *U. maydis* and symbiont *L. bicolor* have a minimal set of PCWDEs what is in accordance with their lifestyle (Fig. 5B; Additional File 9).

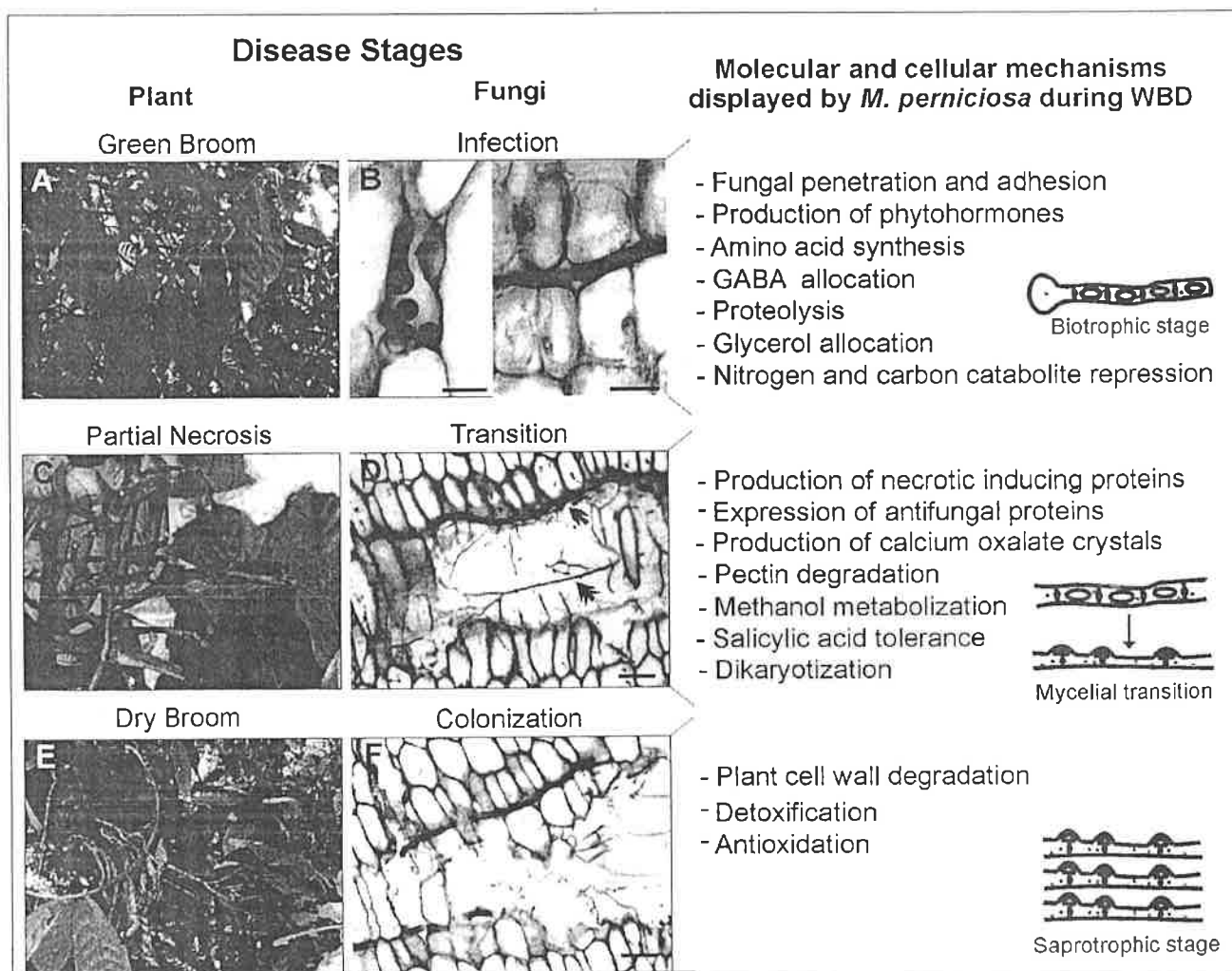
Pectate lyases (PL1 and PL3 according to CAZy nomenclature – <http://www.cazy.org>) cleave pectin, an essential component of plant cell walls. Among the fungi analyzed, *F. graminearum* and *M. perniciosus* contain the largest number of PLs (Fig. 5B; Additional File 9). Unlike *M. grisea*, the other hemibiotrophic analyzed, *F. graminearum* and *M. perniciosus* does not have specialized structures (appressoria) for non-enzymatic penetration of plants, and colonizing the apoplast by breaching the middle lamella barrier (Fig. 7). In addition, both are able to infect dicotyledons that contain cell walls with larger amounts of pectin than monocots [108]. This analysis suggests that PLs have an important role for pathogens that colonize the apoplast of dicotyledons.

Insights into *M. perniciosus* intermediary metabolism and WBD development

Pathogen energy status and the acquisition of host metabolic compounds by the pathogen are factors that determine the outcome of disease onset. Therefore, enzymes of *M. perniciosus* involved in intermediary metabolism (i.e., reactions concerned with storing and generating metabolic energy) may have an important role during WBD (Additional File 3 – Worksheet Functional Annotation). During the first stages of WBD, *M. perniciosus* may encounter a nutrient-poor and extreme oxidative environment containing host defense molecules, such as nitric oxide (NO), an inhibitor of the cytochrome respiratory pathway [109]. Under these conditions, the activity of a NO tolerant alternative oxidase (AOX) may constitute a critical bypass mechanism for the cytochrome pathway [110-112]. A single copy of AOX was detected in *M. perniciosus*, and preliminary experiments indicate that this gene has a higher expression in fungi grown in nutrient-poor media (data not shown).

Following this scenario, in the low-energy status represented by the biotrophic phase, AOX may provide NAD⁺ for the turnover of the mitochondrial tricarboxylic acid cycle (TCA) and the peroxisomal glyoxylate cycle (GLOX). Many reports have documented the importance of GLOX for phytopathogens during host infection [113-115]. All genes coding for proteins of this enzymatic pathway are present in the genome of *M. perniciosus*, including the key enzymes isocitrate lyase and malate synthase (Additional File 3 – Worksheet Functional Annotation).

Curiously, *M. perniciosus* is able to use methanol as the only carbon source (data not shown), indicating that this fungus may have a methylotrophic metabolism. Methylotrophism have been extensively studied in methylotrophic yeasts such as *Pichia angusta* and recently methanol oxidases (MOX) have been described in filamentous fungi such as *C. fulvum* [116] and wood-degrading basidiomycete *Gloeophyllum trabeum* [117]. A gene encoding a MOX was identified in *M. perniciosus* genome (Additional File 3 – Worksheet Functional Annotation). Interestingly, this gene was previously detected as being overexpressed in biotrophic-mycelia [20]. The methanol catabolism enzymes formaldehyde dehydrogenase and formate dehydrogenase (Additional File 3 – Worksheet Functional Annotation) were also found in the genome, which provides evidence that *M. perniciosus* indeed hydrolyzes methanol. Methanol is, possibly, derived from the first step of pectin hydrolyzation performed by the cell wall degrading enzyme pectin methylesterase [118], or from demethylation of lignin that occurs after manganese peroxidase and/or laccase action [117]. We can not discount the possibility that the biotrophic fungi can use the methanol produced by pectin metabolism during normal

**Figure 7**

An illustrated hypothetical model for WBD. Model correlating classical symptoms of green and dry broom in the field (A, C and E), *M. perniciosus* development inside cacao (B: biotrophic stage, D: transition from biotrophic to saprotrophic stage, F: saprotrophic stage) and molecular and cellular events displayed by the fungus in each developmental stages, based on genes annotated in genome survey (right side of the panel). Notice in B (right side) the presence of biotrophic mycelia in the apoplast surrounded by intact living cells. Also notice in C the presence of biotrophic mycelia (arrowhead) and saprotrophic mycelia (double arrowhead) inside a necrotic region. Micrographs scales: B left side: 15 μ M; B right side: 25 μ M; D: 50 μ M; F: 50 μ M.

cell wall synthesis in early stages of plant expansion [119]. The possibility that *M. perniciosus* is methylotrophic resembles the methylbacteria nutritional strategy, which provides an ecological advantage over non-methylotrophic microorganisms present in the phyllosphere [120].

Previous data indicated that the utilization of glycerol, instead of fermentable sugars (i.e., glucose), is an important environmental clue for the maintenance of the biotrophic stage [18,121]. Accordingly, our group detected higher amounts of glycerol during the biotrophic fungal phase of the green brooms development than in dry brooms [13]. Additionally, *in vitro* assays showed that the

shift from glucose to glycerol media increased anti-oxidative defenses of *M. perniciosus* mycelia [122]. This result correlates well with green broom environment (high content of glycerol and ROS) raising the possibility that glycerol is a critical metabolite during the initial stages of the disease cycle. Genes involved in glycerol metabolism and uptake are present in *M. perniciosus*, including a biotrophic induced aquaglyceroporin transporter [20] (Additional File 3 – Worksheet Functional Annotation), suggesting that this fungus is able to acquire extracellular glycerol.

After 35 days of WBD, glucose levels increase again, concomitant with a reduction in starch levels [13]. We found

a gene model similar to a secreted glucoamylase in the *M. perniciosa* genome (Additional File 3 – Worksheet Functional Annotation). Gibberellin is an inducer of α -amylase production [123]. We can envisage a disease scenario in which the fungus produces an extracellular amylase and hormones (i.e., gibberellin) that triggers plant amyolytic activity, which then decrease starch content. The resulting glucose can be utilized by the plant since at this stage the green brooms/infected tissues are rapidly growing or even be acquired by the pathogen at the transition phase between biotrophic and necrotrophic stages.

Nitrogen starvation also appears to be a factor that influences the biotrophic lifestyle [124] and is associated with the expression of pathogenicity genes and PCWDE, mainly in hemibiotrophic and biotrophic fungi that deal with nutrient deprivation during early infection [125]. Our data from microarray and EST analysis suggests that nitrogen catabolite repression (NCR) occurs in *M. perniciosa* by the induction of GABA permease, tRNA synthetases and AROM protein [20]. As reported previously [20] *M. perniciosa* contains a gene homologous to *CLNR1* from the hemibiotrophic fungus *C. lindemuthianum* (Additional File 3 – Worksheet Functional Annotation), a global nitrogen regulator that belongs to the AREA/NIT2 family. *CLNR1* activates enzymes and transporters that enable uptake and catabolism of secondary nitrogen sources [126]. The depletion of *CLNR1* impaired the fungal switch to necrotrophy [124], emphasizing the importance of nitrogen catabolism in hemibiotrophic development. Our genomic data mining identified a gene similar to *NPR2*, which encodes a regulatory protein that may act upstream of the AREA/NIT2 protein (Additional File 3 – Worksheet Functional Annotation). *NPR2* is required for the expression of the *M. grisea* pathogenicity gene *MPG1* [127]. In addition, genes that encode enzymes involved in alternative nitrogen sources uptake (i.e., GABA transporter, urea permease and nitrate transporter) and metabolism (i.e., urease, nitrate reductase, nitrite reductase, arginase and uricase) were identified (Additional File 3 – Worksheet Functional Annotation). The presence of such genes in *M. perniciosa* genome indicates that this fungus could allocate and utilize alternative nitrogen sources in the absence of preferential nitrogen sources (glutamine and ammonia) reinforcing our hypothesis that *M. perniciosa* suffers NCR during early stages of WBD.

Conclusion

Our analysis of the *M. perniciosa* genome survey yielded interesting insights and clues into the molecular mechanisms underlying WBD. As far as we know, this is the first phytopathogen included in the order Agaricales sequenced. Therefore, our results support the investigation of pathogenicity mechanisms among Agaricales and Basidiomycete. In addition, we provide an approach for normalization of gene family data in a genome survey that

can aid the genomics community interested in functional analysis in incomplete genome data.

Based on annotated fungal genes from this report and from previous other studies, we designed a hypothetical model for WBD that correlates plant phenotype changes that happen during the disease with the developmental progression of *M. perniciosa* (Fig. 7).

After penetration and adhesion to the plant, the biotrophic fungus slowly grows inside the apoplast causing a series of phenotype changes in cacao, such as hypertrophy and hyperplasia, phototropism and epinasty, by secreting phytohormones that unbalance cacao metabolism (Fig 7A and 7B). In addition, *M. perniciosa* has to deal with nitrogen deprivation in the apoplast, which signals the production of proteins related to the acquisition of alternative nutrient sources, proteolysis and amino acid synthesis. Furthermore, the fungus takes advantage of the increasing content of glycerol in the green broom, an important cue for biotrophic stage maintenance, by expressing aquaglyceroporins. As previously discussed [13], the starch accumulated in early stages of green brooms seems to be metabolized to glucose, which suggests an amyolytic activity exerted by fungi and/or by plant amylases. This increase of glucose is not accompanied by an increase of fructose, but by a sucrose augmentation. Since photosynthesis is not increased during WBD, we suggest that sucrose is translocated from other tissues to the green brooms. Therefore, we hypothesize that these mechanisms cause a source-to-sink transition in stem, turning green brooms into a drain of nutrients.

After numerous physiological and biochemical changes in the plant, which may be caused by the fungal infection, there occurs a transition from the biotrophic to the saprotrophic lifestyle (Fig. 7C and 7D). This change could result from the increase of nutrients in the fungal environment and may be controlled by an AREA/NIT2-like regulator. During this transition phase, the plant displays the beginning of necrosis at the distal portion of the leaves that could be due to the action of NEPs and cerato-platanins that are expressed in the biotrophic hyphae. In addition, *M. perniciosa* produces PCWDEs, such as pectinases, whose action aids the fungus in breaching the middle lamella barrier. Pectin degradation releases methanol, which in turn could be used by *M. perniciosa* as a carbon source, through the action of a MOX and other methanol metabolizing enzymes. Moreover, the calcium released from pectin disruption could be scavenged by the oxalate synthesized by the fungus, a compound that triggers cell death. The release of cell content during necrosis, and the probable aforementioned source-to-sink transition, may influence, or even be indispensable to the *M. perniciosa* switch from biotrophism to saprotrophism. Thus, the postulated carbon and nitrogen catabolite repression dis-

played by the biotrophic mycelia would be switched off, thus causing the mycelial change to its invasive dikaryotic/saprotrophic stage.

During disease progression, and mainly during colonization of saprotrophic hyphae, *M. pernicioso* must deal with an intense oxidative environment. Based on our analysis this stress can be overcome by the action of several antioxidative and detoxifying enzymes. Furthermore, we believe that *M. pernicioso* exerts a negative control on plant defense against necrotrophic/saprotrophic fungi by producing salicylic acid, which would limit competition by other fungi by the action of antifungal proteins. Finally, after alternating wet and dry periods, the formation of the basidiomes produced by saprotrophic hyphae occurs.

Latin American cacao crops suffer tremendous damages caused by WBD, which mainly affects small acreage farmers. Such impact in the relatively primitive cacao cultivation system not only affects the socio-economic status of farmers but also the preservation of the rainforest. The demand for strategies that limit cacao diseases requires an intense effort in understanding the pathogenicity and plant resistance mechanisms. Further sequencing projects of cacao and its pathogens will serve as a background for the integration of transcriptomics, proteomics and metabolomics of these species in a systems biology approach. Such initiatives will provide tools for biological control, crop management and cacao biotechnology to combat cacao diseases. We believe that our report is the first step towards such an integrative initiative and provides insights into the molecular mechanisms of WBD which can aid the cacao's WBD-concerned community to develop control strategies for this plant-fungus interaction.

Methods

Biological material, libraries construction and sequencing

Total DNA was extracted from saprotrophic hyphae of *M. pernicioso* strain FA553 (CP02) maintained in Malt Yeast Extract Agar (Difco) at 27°C. DNA was extracted from grounded mycelia by incubation in CTAB buffer (CTAB 3%, NaCl 1.4 M, EDTA 20 mM pH 8.0, Tris-HCl 10 mM pH 8.0, PVP 1.0%, β -mercaptoethanol 0.2%) at 65°C during 30 min; followed by one phenol:chloroform:isoamyl alcohol (25/24/1) wash, precipitation with sodium acetate pH 5.2 (0.1 Vol) and cold 100% ethanol (2 Vol). DNA was eluted in deionized water and sheared by nebulization and sonication into fragments of approximately 2 Kbp, which were size selected on agarose gels and purified with S.N.A.P. Gel Purification Kit (Invitrogen – Life technologies, USA). DNA fragments were blunt-end ligated into the pCR4Blunt plasmid (Invitrogen – Life technologies, USA). Approximately 50 genomic libraries were constructed, each one corresponding to individually

growing cultures. Sequencing was done in an ABI Prism 3700 sequencer (Applied Biosystems, USA).

Clustering

The resulting chromatograms were submitted to the *M. pernicioso* database and subjected to automatic base calling using the software PHRED [128]. The contaminating vector sequences and low quality shotgun reads, without at least 100 bp with phred note ≥ 16 were trimmed by using the program LUCY [129]. Shotgun reads showing significant sequence similarity (BLASTn, E-value $\geq 1E-30$) with *M. pernicioso* mitochondrial sequences (see above) were removed from subsequent assemblies. The clustering and assembly were performed using the software PHRAP <http://www.phrap.org>. Afterwards, low quality regions of singlets previously evaluated by LUCY (the last window of 10 bp that has an average probability of error given by phred ≤ 10) were trimmed. The remaining clusters were subjected to similarity searches against the NCBI non-redundant protein and nucleotide database using the BLASTx and tBLASTx, respectively, with an E-value cutoff of 1E-5.

Genome Length Statistical Validations

Statistical analyses of the genome length were performed using two approaches. The first one was based on the Dog genome survey using counting of start positions offsets for overlapping reads [31] (for further details see Additional File 1). The second was based on Lander Waterman Theory, [34] and their applications [35,36], which estimate the theoretical values of expected number of clusters (contigs + singlets), contigs, gaps, average cluster size and average gap size, using the effective average read length (L), the total number of reads in the assembly (N) and the estimated genome size (G). L is the average number of base pairs of a read that contributes to the contig through parsing of ace file <http://www.phrap.org>. In our analysis L was equal to 550 bp. In order to estimate the gap size distribution in the *M. pernicioso* genome survey, we performed a comparison between a set of eukaryotic core protein (generated by CEGMA pipeline [37]) and *M. pernicioso* contigs using TBLASTN with threshold of 1e-10 for the E-value (Further details in Additional File 1). The estimation of misassembled sequences due to repetitive regions in the genome was performed using the integrated pipeline amosvalidate [38] (Further details in Additional File 1).

Gene Finding

EST against genome alignments

The alignment of ESTs with genomic sequences was performed using the package GeneSeqer [130] with the pre-built *Aspergillus* intron model. 300 highly confident introns were selected and used as an input for Exalin program [39] that is able to build a splice site model for an organism. The positions of the splice sites as assigned by

Exalin were used to rank overlapping gene predictions (see below).

Ab initio gene models prediction

The *ab initio* gene models prediction was performed with the trainable, open source gene predictors AUGUSTUS [40], SNAP [41] and GENEZILLA [42]. Ten copies of an artificial sequence of 240 Kbp (total of 2.4 Mbp) formed by the concatenation of the *M. perniciosus* ESTs coding regions, together with a *C. cinerea* gene dataset containing 1.2 Mbp were submitted to "pre-training" in AUGUSTUS gene predictor. *M. perniciosus* resulting predictions were compared with the protein databank NR using BLASTp. The predictions with similarities in NR, and with coverage $\geq 90\%$, were selected. After redundancy elimination, *M. perniciosus* gene models were used to train the three gene finders aforementioned. Predictions with less than 30 amino acids were eliminated, and the remaining predictions were grouped in overlapping clusters.

Ab initio gene models ranking

The predictions in each overlapping cluster were ranked according to the criteria used by the Fungal Genome Initiative at Broad institute <http://www.broad.mit.edu>. In each cluster, the "best" *ab initio* gene model according to the stipulated criteria was selected for functional annotation. The criteria for the ranking of the gene models were the following:

1. Manual annotation had priority over all other evidences;
2. Predictions with EST evidences had priority over the predictions without EST evidences;
3. If two predictions had EST evidences, the one with more splice sites in exact agreement with ESTs had priority;
4. Prediction with similarities with known proteins had priority. A prediction was considered to be similar to some known protein if it had an E-value of at most $1e-10$ (BLASTP against NR+*Phanerochaete chrysosporium* protein set);
5. If two predictions had similarity with known proteins, the one with better coverage score had priority. The coverage score was defined as $2 \times CP \times CH / (CP + CH)$, where CP is the coverage of the prediction and CH is the coverage of the similar protein;
6. In clusters without similarity with known proteins and without EST evidence, the priority was for AUGUSTUS, SNAP and GENEZILLA, in this order. This criterion was chosen according to the performance of the three pro-

grams in a dataset of 60 genes structures visually inspected.

A final filter discarded gene predictions reported by only one program, without similarity to known proteins and without EST evidence.

Extrinsic gene models prediction

The extrinsic gene model predictions were performed by two methodologies. First, 17,991 contigs and 7,065 singletons were submitted to similarity analysis in a databank containing BLASTX-NR plus *Phanerochaete chrysosporium* proteins. The genomic regions containing homologues in this databank were selected and assigned as putative gene models. GenomeThreader [131] program was used to make protein-DNA spliced alignments between the BLAST first hit against and the genomic sequence, serving as a guide to delimit the start and stop codons and exon-intron boundaries of the regions of the contigs containing similarity with GenBank.

Concurrently, *M. perniciosus* ESTs aligned with genomic clusters (see above) were inspected to verify if the region in which they aligned contained a BLAST extrinsic prediction gene model. These extrinsic gene models (EST and BLAST) were compared with each other to evaluate the amount of gene models predicted by these methods. After these comparisons, the extrinsic gene models were divided into 4 datasets:

- (i) ESTMODELS: retrieved from the spliced alignments of the ESTs against the genomic clusters not covered to a BLAST extrinsic gene models prediction. Low score spliced alignments and ESTs that seen clearly to be UTR of a neighboring prediction were not included.
- (ii) BLASTMODELS: derived from BLAST extrinsic gene models predictions analysis covering genomic regions without EST evidence.
- (iii) COMBINEDMODELS: gene models derived from genomic sequence regions with BLAST hit and EST evidence.
- (iv) CURATEDMODELS: extrinsic predictions manually annotated for manual correction of merged or split predictions. The genomic survey and gene models nomenclature are depicted in additional file 10.

tRNA prediction

For tRNA prediction, the tRNAscan-SE program [132] was taken into account with the default parameters, which searched for conserved sequences and the characteristic secondary structure of tRNAs.

MCL-families clustering

MCL graph clustering algorithm was applied to generate *M. perniciosus* gene families using WU-TBLASTx "all against all" as the tool used for aligning the gene models [43].

Normalization of gene family data

The normalization of gene family data was performed by comparison between *C. cinerea* proteins and *M. perniciosus* gene models, using TBLASTN with 1E-10 of E-value threshold. The number of gene models similar to a *C. cinerea* protein was plotted according to protein length. The equation that estimates the number of gene models representing the same protein was generated using linear regression fitting. This equation was used to estimate a normalization factor to each MCL family and CDD-PFAM domain according to the average of protein length of their members.

Automatic Annotation and Metabolic Maps

The automatic annotation program AutoFACT [47] was used for functional annotation of gene models. The set of coding sequences from gene models were submitted to similarity searches against the UNIREF100, UNIREF90, NR, and KEGG databases using BLASTx (E-value $\leq 1E-5$) and against CDD-PFAM using RPS-BLAST (E-value $\leq 1E-5$) [48]. These results were submitted to AutoFACT, which searches for a consensus in the results and output descriptions and statistics about protein domains and families.

For an inference of *M. perniciosus* metabolic maps, we used Pathway Tools (version 11.0), a software of BioCyc databases [49], which generates a metabolic map from a previously annotated genome. The pathways that are probably present in the genome are imported from a reference database, following the Pathway Tools parameters [133]. The annotated genome input was obtained from EST manual annotation and from *M. perniciosus* gene models AutoFACT annotation, using as main information the product name and, if available, E.C. numbers. Metabolic pathways of interest were manually annotated for the elimination of false positives.

The analysis of transporters was made based on a BLASTX search of *M. perniciosus* gene models and other fungal genes against TCDB (Transport Classification Database - <http://www.tcdb.org>), using a threshold of E-value 1E-05. All classes from third level that contained at least one species with 2% or more of representations were separately represented in the results.

Selection and expression confirmation of no hits *ab initio* gene models

The selection of *ab initio* gene models was performed using a SQL query wizard. As input, we ask for gene models without similarity in GenBank NR, which encoded

proteins that contained at least 2 cysteines and a signal peptide, previously identified by Signal-P 3.0 program [134]. 74 gene models were then selected and their nucleotide sequences were used as template for the design of primers nested in: (i) the sequence encoding the putative signal peptide (SPE); (ii) the sequence encoding the putative first amino acid of mature protein (MAT) and; (iii) the sequence containing the putative stop codon (END). The latter was designed in reverse complement ("reverse") to allow gene amplification using the other two primers ("forward"). RT-PCR analysis was performed to validate the expression of *ab initio* predicted gene models. RNA from saprotrophic mycelia was extracted using hot-phenol method with modifications [135]. Equal amounts of total RNA from CP02 saprotrophic mycelia cultures (24 h, 48 h, 4 days and 7 days) were mixed. After DNase (Invitrogen, USA) treatment, 2 μ g of total RNA was reverse transcribed using Superscript RTIII (Invitrogen, USA) in a total volume of 20 μ L, following the manufacturer's instructions. PCR reactions were conducted according to primers (MWG, Imprint Genetics Corp) temperature of melting (TMs).

Abbreviations

ABC: ATP Binding Cassette; AOX: Alternative oxidase; bp: base pairs; COC: calcium oxalate crystals; CTAB: Cetyl trimethylammonium bromide; EDTA: Ethylenediaminetetraacetic acid; EST: Expressed Sequence Tag; FPR: Frosty Pod Rot; GA: Gibberellin; GABA: Gamma-aminobutyric acid; GGPP: Geranylgeranyl Diphosphate; IAA: indole-3-acetic acid; JA: Jasmonic acid; Kbp: One thousand base pairs; LW: Lander Waterman; Mbp: One million base pairs; MOX: Methanol oxidase; NAD(P): Nicotinamide adenine dinucleotide phosphate; NCR: nitrogen catabolite repression; NEP: Necrosis and ethylene-inducing proteins; NO: Nitric oxide; PCD: Programmed cell death; PCWDE: Plant cell wall degrading enzymes; PVP: Poly(vinylpyrrolidone); ROS: Reactive oxygen species; RT-PCR: Reverse Transcription-Polymerase Chain Reaction; SA: Salicylic acid; PCR: Polymerase Chain Reaction; PR: Pathogenesis related protein; TCA: Tricarboxylic acid cycle; TE: Transposable Elements; TM: temperature of melting; tRNA: transfer RNA; UTR: untranslated region; WBD: Witches' Broom Disease.

Authors' contributions

JMCM conceived and wrote the article, was responsible for gene models annotation, gene density analysis, primer design and data interpretation. MFC conceived bioinformatics analysis and was responsible for genome assembly, BLASTs, statistical analysis and data interpretation. GGLC was responsible for gene models predictions, Genome Threader analysis, MCL clustering analysis and gene browser design and implementation. EFF participated in genome assembly, was responsible for transporters analy-

sis and annotation, and made figures edition. LPP was responsible for AutoFACT and BioCyc implementation and analysis. JR was responsible for EST library construction and sequencing and participated in genome annotation. CC, DMC, AFC and HC were responsible for *M. perniciosus* DNA extraction, genome shotgun, cloning and maintenance of genomic libraries. ROV was responsible for gene browser design and implementation, tRNA analysis and *M. perniciosus* sketches. RCE performed RT-PCR and aided in primer design. OG, DPTT BVO and MHM participated in genome annotation and were responsible for the characterization of genes encoding NEPs, AOX, MOX and auxin biosynthetic proteins, respectively. ABLP was responsible for EST library construction. MCSR was responsible for microscopic analysis. MRAR participated in genome assembly and in the development of bioinformatics tools. IABC, KPG, MSG, AGN and LBB were coordinators of sequencing groups. JPMN was responsible for sequencing and libraries maintenance. MJG and BAB participated in final manuscript elaboration and provided additional sequencing in order to close gene gaps. LWM participated in pathogenicity gene data interpretation and final manuscript elaboration. JCC was the co-coordinator of *M. perniciosus* genome project and coordinated UESC sequencing group. GAGP conceived and was the coordinator of *M. perniciosus* genome project, participated in data interpretation and in final manuscript elaboration.

Additional material

Additional file 1

Genome statistical validations. A) Estimation of genome length using dog genome survey protocol, B) Estimate of distribution of gap sizes in *M. perniciosus* genome assembly, C) Estimate of misassembly sequences due to repetitive regions.

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2164-9-548-S1.pdf>]

Additional file 2

MCL analysis of *M. perniciosus* gene models. All predicted proteins were compared all-against-all using WU-TBLASTX. A score ($-\log(E\text{-value})$) for each pair of proteins (u, v) with significant BLAST hits (Evalue $1e-5$) was assigned. The MCL algorithm (inflation parameter 2.0) was applied to find clusters in this graph. This method is fully automatic and protein clusters reported were not subjected to manual curation. ID: number of the MCL family; #members: number of gene models present in each family; Norm factor: factor used to normalize the number of gene models present in each family (see methods); Norm#members: normalized number of gene models present in each family. Annotation: words associated to each family after correlation of gene models with AutoFACT annotation. In parenthesis are the occurrence numbers of each word. Each worksheet shows the ranking of families using normalization factor (Rank_Norm) or not using this factor (Rank_Non_Norm).

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2164-9-548-S2.xls>]

Additional file 3

Functional annotation of *M. perniciosus* gene models discussed in this paper. ID: gene model; First Hit (BLASTX-NR): Most similar sequence in GenBank; E-value: E-value of most similar sequence; AutoFACT annotation: automatic annotation by AutoFACT; AutoFACT E-value: E-value of AutoFACT annotation; EST: presence (Y) or absence of an EST aligned in this gene model; MCL family: family annotated by MCL analysis.

Worksheet P450: annotation of gene models similar to cytochrome P450 monooxygenases; Worksheet transposons: classification and annotation of gene models similar to transposable elements; Worksheet unknown gene families; annotation of top 20 MCL unknown gene families; Worksheet functional annotation: classification and annotation of gene models similar to efflux transporters, anti-oxidative enzymes, phytohormones biosynthesis related proteins, pheromone receptors, salicylate hydroxylases, effectors/elicitors/pathogenicity associated proteins, cell wall degrading enzymes and intermediary metabolism enzymes (cytochrome pathway bypass, Glyoxylate pathway and oxalate formation, glycerol uptake and metabolism, extracellular sugar degrading enzymes and nitrogen regulation, uptake and metabolism enzymes), EC - enzyme classification (<http://expasy.org/enzyme/>); Worksheet transporters; Relative percentage of transporters distribution in fungi genomes (see methods).

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2164-9-548-S3.xls>]

Additional file 4

Ranking of CDD-PFAM families annotated in *M. perniciosus*. Gene models were annotated based on CDD-PFAM-ID and ranked. This analysis was performed with other fungi genomes, which CDD-PFAM entries were classified according to *M. perniciosus* ranking. CDD-ID: CDD entry; PFAM Domain: PFAM entry; #Hits Domains: number of gene models containing each CDD-PFAM domain; %Hits Domains: percentage of gene models containing each CDD-PFAM domain in relation to total number of gene models containing a CDD-PFAM domain; %Hits PTN: Percentage of gene models containing each CDD-PFAM domain in relation to total number of *M. perniciosus* gene models; Rank: non-normalized *M. perniciosus* CDD-PFAM ranking; PTNS: proteins in each organism; Norm Factor: factor used to normalize the number of gene models containing each CDD-PFAM domain; # Hits Domains Norm: normalized number of gene models containing a CDD-PFAM domain; %Hits Domains Norm: percentage of gene models containing each CDD-PFAM domain in relation to total number of gene models containing each CDD-PFAM domain; %Hits PTN Norm: Normalized percentage of gene models containing each CDD-PFAM domain in relation to total number of *M. perniciosus* gene models; Rank Norm: normalized *M. perniciosus* CDD-PFAM ranking. Worksheets show the ranking of CDD-PFAM domains using normalization (Rank_Norm) or not using normalization (Rank_Non_Norm).

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2164-9-548-S4.xls>]

Additional file 5

BioCyc comparison between *S. cerevisiae* and *M. perniciosus* metabolic pathways. Worksheet Compounds: Comparison of number of reactions in each organism containing the compounds described in the table; Worksheet pathways: Comparison of number of pathways in each organism present in each pathway class. The two largest top-level classes, Biosynthesis and Degradation/Utilization/Assimilation, are broken down further to show the distribution of pathways among their next-level subclasses.

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2164-9-548-S5.xls>]

Additional file 6

Annotation of gene models with no similarity in BLASTX-NR encoding hypothetical small secreted proteins containing at least 2 cysteines. ID: gene model; # residuos: number of amino acids of predicted protein encoded by the gene model; # cysteines: number of cysteines in predicted protein; Binomial RT; statistical analysis of cysteines presence in gene models (see methodological details in the file).

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2164-9-548-S6.xls>]

Additional file 7

Primers used in the amplification of no hits gene models encoding hypothetical small secreted proteins containing at least 2 cysteines. ID: gene model; Set of primers: Group of primers used for the amplification of a gene model. Primer sequence: sequence of primers (SPE – nested in sequence encoding the putative signal peptide; MAT – nested in sequence encoding the putative first amino acid of mature protein; END – nested in sequence containing the putative stop codon). Amplification: positive (Y) or negative (N); EST: presence (Y) or absence (N) of an EST aligned in this gene model; Length: length of amplicon in genomic and cDNA using two combinations of primers (SPE-END; MAT-END).

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2164-9-548-S7.xls>]

Additional file 8

Examples of amplifications of no hits gene models. PCR amplicons were run on 1% agarose gels. SPE: amplicons resulted from amplification with SPE and END primers; MAT: amplicons resulted from amplification with MAT and END primers; Ctl: water as template (control); Gen: genomic DNA as template; Glu: cDNA from saprotrophic mycelia grown in glucose as template; Cac: cDNA from saprotrophic mycelia grown in cacao extract as template; M: DNA molecular marker.

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2164-9-548-S8.pdf>]

Additional file 9

Comparison of plant cell wall degrading enzymes in fungi that interact with plants. PFAM entries were correlated with the CAZy nomenclature <http://www.cazy.org> of plant cell wall degrading enzymes.

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2164-9-548-S9.pdf>]

Additional file 10

Genomic survey sequences and gene models nomenclature.

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2164-9-548-S10.pdf>]

Acknowledgements

We thank CENAPAD-SP (Centro Nacional de Processamento de Alto Desempenho de São Paulo) for data processing support, all sequencing technician for excellent technical assistance, especially Robson Dias (UESC – Brazil), Sulamita Freitas Franco (UNICAMP – Brazil) and Vitor Hugo de Almeida e Silva (UNICAMP – Brazil), and Dr. Francisco Javier Medrano for review and commentaries about the article. We are indebted to CNPq (Conselho Nacional de Desenvolvimento Científico e Tecnológico,

Projects 68.0032/2001-0 and 42.2279/2006-8), SEAGRI (Secretaria de Agricultura, Irrigação e Reforma Agrária do Estado da Bahia), CARGILL/FUNCAMP/UNICAMP convenio (number C. 3007), FAPESP (Fundação de Amparo à Pesquisa do Estado de São Paulo) and CAPES (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior) for financial support.

References

- Purdy LH, Schmidt RA: **Status of cacao witches' broom: Biology, epidemiology, and management.** *Annu Rev Phytopathol* 1996, **34**:573-594.
- Aime MC, Phillips-Mora W: **The causal agents of witches' broom and frosty pod rot of cacao (chocolate, *Theobroma cacao*) form a new lineage of Marasmiaceae.** *Mycologia* 2005, **97**(5):1012-1022.
- Griffith GW, Nicholson J, Nenninger A, Birch RN, Hedger JN: **Witches' brooms and frosty pods: two major pathogens of cacao.** *New Zeal J Bot* 2003, **41**(3):423-435.
- Evans HC: **Cacao diseases – The trilogy revisited.** *Phytopathology* 2007, **97**(12):1640-1643.
- Pereira JL, Ram A, Figueiredo JM, de Almeida LC: **La primera aparición de la "escoba de bruja" en la principal área productora de cacao del Brasil.** *Turrialba* 1989, **39**(7):459-461.
- Pereira JL, de Almeida LCC, Santos SM: **Witches' broom disease of cocoa in Bahia: Attempts at eradication and containment.** *Crop Protection* 1996, **15**(8):743-752.
- Evans HC: **Pleomorphism in *Crinipellis pernicioso*, Causal Agent of Witches Broom Disease of Cocoa.** *Trans Br Mycol Soc* 1980, **74**(Jun):515-523.
- Griffith GW, Hedger JN: **Dual culture of *Crinipellis pernicioso* and potato callus.** *Eur J Plant Pathol* 1994, **100**(6):371-379.
- Delgado JC, Cook AA: **Nuclear condition of basidia, basidiospores, and mycelium of *Marasmius perniciosus*.** *Can J Bot* 1976, **54**(1-2):66-72.
- Rincones J, Meinhardt LW, Vidal BC, Pereira GA: **Electrophoretic karyotype analysis of *Crinipellis pernicioso*, the causal agent of witches' broom disease of *Theobroma cacao*.** *Mycol Res* 2003, **107**(Pt 4):452-458.
- Rincones J, Mazotti GD, Griffith GW, Pomela A, Figueira A, Leal GA Jr, Queiroz MV, Pereira JF, Azevedo RA, Pereira GA, Meinhardt LW: **Genetic variability and chromosome-length polymorphisms of the witches' broom pathogen *Crinipellis pernicioso* from various plant hosts in South America.** *Mycol Res* 2006, **110**(Pt 7):821-832.
- Formighieri EF, Tiburcio RA, Armas ED, Medrano FJ, Shimo H, Carels N, Góes-Neto A, Cotomacci C, Carazzolle MF, Sardinha-Pinto N, Rincones J, Digiamietri L, Carraro DM, Azeredo-Espin AM, Reis SF, Deckmann AC, Gramacho K, Gonçalves MS, Moura Neto JP, Barbosa LV, Meinhardt LW, Cascardo JCM, Pereira GAG: **The mitochondrial genome of the phytopathogenic basidiomycete *Moniliophthora pernicioso* is 109 kb in size and contains a stable integrated plasmid.** *Mycol Res* 2008, **112**:1136-1152.
- Scarpini LM, Meinhardt LW, Mazzafera P, Pomella AW, Schiavinato MA, Cascardo JC, Pereira GA: **Biochemical changes during the development of witches' broom: the most important disease of cocoa in Brazil caused by *Crinipellis pernicioso*.** *J Exp Bot* 2005, **56**(413):865-877.
- Gesteira Ada S, Micheli F, Ferreira CF, Cascardo JC: **Isolation and purification of functional total RNA from different organs of cacao tree during its interaction with the pathogen *Crinipellis pernicioso*.** *Biotechniques* 2003, **35**(3):494-496. 498-500
- Pirovani CP, Carvalho HA, Machado RC, Gomes DS, Alvim FC, Pomella AV, Gramacho KP, Cascardo JC, Pereira GA, Micheli F: **Protein extraction for proteome analysis from cacao leaves and meristems, organs infected by *Moniliophthora pernicioso*, the causal agent of the witches' broom disease.** *Electrophoresis* 2008, **29**(11):2391-2401.
- Lima JO, dos Santos JK, Pereira JF, de Resende ML, de Araujo EF, de Queiroz MV: **Development of a transformation system for *Crinipellis pernicioso*, the causal agent of witches' broom in cacao plants.** *Curr Genet* 2003, **42**(4):236-240.
- Filho DF, Pungartnik C, Cascardo JC, Brendel M: **Broken hyphae of the basidiomycete *Crinipellis pernicioso* allow quantitative assay of toxicity.** *Curr Microbiol* 2006, **52**(5):407-412.
- Meinhardt LW, Bellato CM, Rincones J, Azevedo RA, Cascardo JC, Pereira GA: **In vitro production of biotrophic-like cultures of**

- Crinipellis pernicioso*, the causal agent of witches' broom disease of *Theobroma cacao*. *Curr Microbiol* 2006, **52**(3):191-196.
19. Garcia O, Macedo JA, Tiburcio R, Zapparoli G, Rincones J, Bittencourt LM, Ceita GO, Micheli F, Gesteira A, Mariano AC, Schiavinato MA, Medrano FJ, Meinhardt LW, Pereira GA, Cascardo JC: **Characterization of necrosis and ethylene-inducing proteins (NEP) in the basidiomycete *Moniliophthora pernicioso*, the causal agent of witches' broom in *Theobroma cacao***. *Mycol Res* 2007, **111**(Pt 4):443-455.
 20. Rincones J, Scarpari LM, Carazzolle MF, Mondego JMC, Formighieri EF, Barrau JG, Costa GGL, Carraro DM, Brentani HP, Vilas-Boas LA, Oliveira BV, Sabha M, Dias R, Cascardo JM, Azevedo RA, Meinhardt LW, Pereira GAG: **Differential gene expression between the biotrophic-like and saprotrophic mycelia of the Witches' broom pathogen *Moniliophthora pernicioso***. *Mol Plant Microbe Interact* 2008, **21**(7):891-908.
 21. Gesteira AS, Micheli F, Carels N, Da Silva AC, Gramacho KP, Schuster I, Macedo JN, Pereira GA, Cascardo JC: **Comparative Analysis of Expressed Genes from Cacao Meristems Infected by *Moniliophthora pernicioso***. *Ann Bot (Lond)* 2007, **100**(1):129-140.
 22. Leal GA, Albuquerque PSB, Figueira A: **Genes differentially expressed in *Theobroma cacao* associated with resistance to witches' broom disease caused by *Crinipellis pernicioso***. *Mol Plant Pathol* 2007, **8**(3):279-292.
 23. Munch S, Lingner U, Floss DS, Ludwig N, Sauer N, Deising HB: **The hemibiotrophic lifestyle of *Colletotrichum* species**. *J Plant Physiol* 2008, **165**(1):41-51.
 24. Meinhardt LW, Rincones J, Bailey BA, Aime MC, Griffith GW, Zhang D, Pereira GA: ***Moniliophthora pernicioso*, the causal agent of witches' broom disease of cacao: what's new from this old foe?** *Mol Plant Pathol* 2008, **9**(5):577-588.
 25. James TY, Kauff F, Schoch CL, Matheny PB, Hofstetter V, Cox CJ, Celio G, Guaidan C, Fraker E, Miadlikowska J, Lumbsch HT, Rauhut A, Reeb V, Arnold AE, Amtoft A, Stajich JE, Hosaka K, Sung GH, Johnson D, O'Rourke B, Crockett M, Binder M, Curtis JM, Slot JC, Wang Z, Wilson AW, Schussler A, Longcore JE, O'Donnell K, Mozley-Standridge S, Porter D, Letcher PM, Powell MJ, Taylor JW, White MM, Griffith GW, Davies DR, Humber RA, Morton JB, Sugiyama J, Rossman AY, Rogers JD, Pfister DH, Hewitt D, Hansen K, Hambleton S, Shoemaker RA, Kohlmeyer J, Volkmann-Kohlmeier B, Spotts RA, Serdani M, Crous PW, Hughes KW, Matsuura K, Langer E, Langer G, Untereiner WA, Lucking R, Budel B, Geiser DM, Aptroot A, Diederich P, Schmitt I, Schultz M, Yahr R, Hibbett DS, Lutzoni F, McLaughlin DJ, Spatafora JW, Vilgalys R: **Reconstructing the early evolution of Fungi using a six-gene phylogeny**. *Nature* 2006, **443**(7113):818-822.
 26. Bouck J, Miller W, Gorrell JH, Muzny D, Gibbs RA: **Analysis of the quality and utility of random shotgun sequencing at low redundancies**. *Genome Res* 1998, **8**(10):1074-1084.
 27. Branscomb E, Predki P: **On the high value of low standards**. *J Bacteriol* 2002, **184**(23):6406-6409. discussion 6409.
 28. Fraser CM, Eisen JA, Nelson KE, Paulsen IT, Salzberg SL: **The value of complete microbial genome sequencing (you get what you pay for)**. *J Bacteriol* 2002, **184**(23):6403-6405. discussion 6405.
 29. Gardner SN, Lam MW, Smith JR, Torres CL, Slezak TR: **Draft versus finished sequence data for DNA and protein diagnostic signature development**. *Nucleic Acids Res* 2005, **33**(18):5838-5850.
 30. Taudien S, Ebersberger I, Glockner G, Platzer M: **Should the draft chimpanzee sequence be finished?** *Trends Genet* 2006, **22**(3):122-125.
 31. Kirkness EF, Bafna V, Halpern AL, Levy S, Remington K, Rusch DB, Delcher AL, Pop M, Wang W, Fraser CM, Venter JC: **The dog genome: survey sequencing and comparative analysis**. *Science* 2003, **301**(5641):1898-1903.
 32. Green P: **2x genomes does depth matter?** *Genome Res* 2007, **17**(11):1547-1549.
 33. Woolfit M, Rozpedowska E, Piskur J, Wolfe KH: **Genome survey sequencing of the wine spoilage yeast *Dekkera (Brettanomyces) bruxellensis***. *Eukaryot Cell* 2007, **6**(4):721-733.
 34. Lander ES, Waterman MS: **Genomic mapping by fingerprinting random clones: a mathematical analysis**. *Genomics* 1988, **2**(3):231-239.
 35. Wendl MC, Waterston RH: **Generalized gap model for bacterial artificial chromosome clone fingerprint mapping and shotgun sequencing**. *Genome Res* 2002, **12**(12):1943-1949.
 36. Wendl MC, Yang SP: **Gap statistics for whole genome shotgun DNA sequencing projects**. *Bioinformatics* 2004, **20**(10):1527-1534.
 37. Parra G, Bradnam K, Korf I: **CEGMA: a pipeline to accurately annotate core genes in eukaryotic genomes**. *Bioinformatics* 2007, **23**(9):1061-1067.
 38. Phillippy AM, Schatz MC, Pop M: **Genome assembly forensics: finding the elusive mis-assembly**. *Genome Biol* 2008, **9**(3):R55.
 39. Zhang M, Gish W: **Improved spliced alignment from an information theoretic approach**. *Bioinformatics* 2006, **22**(1):13-20.
 40. Stanke M, Steinkamp R, Waack S, Morgenstern B: **AUGUSTUS: a web server for gene finding in eukaryotes**. *Nucleic Acids Res* 2004:V309-312.
 41. Korf I: **Gene finding in novel genomes**. *BMC Bioinformatics* 2004, **5**:59.
 42. Allen JE, Majoros WH, Pertea M, Salzberg SL: **JIGSAW, GeneZilla, and GlimmerHMM: puzzling out the features of human genes in the ENCODE regions**. *Genome Biol* 2006, **7**(Suppl 1):S91-133.
 43. Enright AJ, van Dongen S, Ouzounis CA: **An efficient algorithm for large-scale detection of protein families**. *Nucleic Acids Res* 2002, **30**(7):1575-1584.
 44. Enright AJ, Kunin V, Ouzounis CA: **Protein families and TRIBES in genome sequence space**. *Nucleic Acids Res* 2003, **31**(15):4632-4638.
 45. Tekaia F, Latge JP: ***Aspergillus fumigatus*: saprophyte or pathogen?** *Curr Opin Microbiol* 2005, **8**(4):385-392.
 46. Kupfer DM, Reece CA, Clifton SW, Roe BA, Prade RA: **Multicellular ascomycetous fungal genomes contain more than 8000 genes**. *Fungal Genet Biol* 1997, **21**(3):364-372.
 47. Koski LB, Gray MW, Lang BF, Burger G: **AutoFACT: an automatic functional annotation and classification tool**. *BMC Bioinformatics* 2005, **6**:151.
 48. Marchler-Bauer A, Anderson JB, Derbyshire MK, DeWeese-Scott C, Gonzales NR, Gwladz M, Hao LN, He SQ, Hurwitz DI, Jackson JD, Ke ZX, Krylov D, Lanczycki CJ, Liebert CA, Liu CL, Lu F, Lu SN, Marchler GH, Mullokandov M, Song JS, Thanki N, Yamashita RA, Yin JJ, Zhang DC, Bryant SH: **CDD: a conserved domain database for interactive domain family analysis**. *Nucleic Acids Res* 2007, **35**:D237-D240.
 49. Karp PD, Ouzounis CA, Moore-Kochlacs C, Goldovsky L, Kaipa P, Ahren D, Tsoka S, Darzentas N, Kunin V, Lopez-Bigas N: **Expansion of the BioCyc collection of pathway/genome databases to 160 genomes**. *Nucleic Acids Res* 2005, **33**(19):6083-6089.
 50. Gonzalez FJ, Nebert DW: **Evolution of the P450 gene superfamily: animal-plant 'warfare', molecular drive and human genetic differences in drug oxidation**. *Trends Genet* 1990, **6**(6):182-186.
 51. Saier MH Jr, Tran CV, Barabote RD: **TCDB: the Transporter Classification Database for membrane transport protein analyses and information**. *Nucleic Acids Res* 2006:D181-186.
 52. Stergiopoulos I, Zwiers LH, De Waard MA: **The ABC transporter MgAtr4 is a virulence factor of *Mycosphaerella graminicola* that affects colonization of subterranean cavities in wheat leaves**. *Mol Plant Microbe Interact* 2003, **16**(8):689-698.
 53. Rogers B, Decottignies A, Kolaczowski M, Carvajal E, Balzi E, Goffeau A: **The pleiotropic drug ABC transporters from *Saccharomyces cerevisiae***. *J Mol Microbiol Biotechnol* 2001, **3**(2):207-214.
 54. Benito B, Garcadeblas B, Rodriguez-Navarro A: **Potassium- or sodium-efflux ATPase, a key enzyme in the evolution of fungi**. *Microbiology* 2002, **148**(Pt 4):933-941.
 55. Apel K, Hirt H: **Reactive oxygen species: metabolism, oxidative stress, and signal transduction**. *Annu Rev Plant Biol* 2004, **55**:373-399.
 56. Able AJ: **Role of reactive oxygen species in the response of barley to necrotrophic pathogens**. *Protoplasma* 2003, **221**(1-2):137-143.
 57. Govrin EM, Levine A: **The hypersensitive response facilitates plant infection by the necrotrophic pathogen *Botrytis cinerea***. *Curr Biol* 2000, **10**(13):751-757.
 58. Rio MCS, Oliveira BV, Thomazella DP, Fracassi da Silva JA, Pereira GAG: **Production of calcium oxalate crystals by the basidiomycete *Moniliophthora pernicioso*, the causal agent of the Witches' Broom Disease of cacao**. *Curr Microbiol* 2008, **56**(4):363-370.

59. Ceita GO, Macêdo JNA, Santos TB, Alemanno L, Gesteira AS, Micheli F, Mariano AC, Gramacho KP, Silva DC, Meinhardt LW, Mazzafera P, Pereira GAG, Cascardo JM: **Involvement of calcium oxalate degradation during programmed cell death in *Theobroma cacao* tissues triggered by the hemibiotrophic fungus *Moniliophthora perniciosa*.** *Plant Sci* 2007, **173**(2):106-117.
60. Garcia-Brugger A, Lamotte O, Vandelle E, Bourque S, Lecourieux D, Poinssot B, Wendehenne D, Pugin A: **Early signaling events induced by elicitors of plant defenses.** *Mol Plant Microbe Interact* 2006, **19**(7):711-724.
61. Kim KS, Min JY, Dickman MB: **Oxalic Acid Is an Elicitor of Plant Programmed Cell Death during *Sclerotinia sclerotiorum* Disease Development.** *Mol Plant Microbe Interact* 2008, **21**(5):605-612.
62. Bolker M: ***Ustilago maydis*-a valuable model system for the study of fungal dimorphism and virulence.** *Microbiology* 2001, **147**(Pt 6):1395-1401.
63. de Arruda MCCR, Ferreira MASV, Felipe MSS: **Comparison of *Crinipellis perniciosa* isolates from Brazil by ERIC repetitive element sequence-based PCR genomic fingerprinting.** *Plant Pathology* 2003, **52**(2):236-244.
64. Andebrhan T, Figueira A, Yamada MM, Cascardo J, Furtek DB: **Molecular fingerprinting suggests two primary outbreaks of witches' broom disease (*Crinipellis perniciosa*) of *Theobroma cacao* in Bahia, Brazil.** *Eur J Plant Pathol* 1999, **105**(2):167-175.
65. Bartley BGD: **Cacao (*Theobroma cacao* L.). Breeding for Durable Resistance in Perennial Crops.** *FAO Plant Production and Protection Paper* 1986, **70**:25-42.
66. Larrondo LF, Canessa P, Vicuna R, Stewart P, Wymelenberg A, Vanden, Cullen D: **Structure and transcriptional impact of divergent repetitive elements inserted within *Phanerochaete chrysosporium* strain RP-78 genes.** *Mol Genet Genomics* 2007, **277**(1):43-55.
67. Poulter RT, Goodwin TJ: **DIRS-I and the other tyrosine recombinase retrotransposons.** *Cytogenet Genome Res* 2005, **110**(1-4):575-588.
68. DeMarco R, Venancio TM, Verjovski-Almeida S: **SmTRC1, a novel *Schistosoma mansoni* DNA transposon, discloses new families of animal and fungi transposons belonging to the CACTA superfamily.** *BMC Evol Biol* 2006, **6**:89.
69. Rep M, Does HC van der, Cornelissen BJ: **Drifter, a novel, low copy hATlike transposon in *Fusarium oxysporum* is activated during starvation.** *Fungal Genet Biol* 2005, **42**(6):546-553.
70. Kempken F, Kuck U: **restless, an active Ac-like transposon from the fungus *Tolypocladium inflatum*: structure, expression, and alternative RNA splicing.** *Mol Cell Biol* 1996, **16**(11):6563-6572.
71. Monroy F, Sheppard DC: **Taf1: a class II transposon of *Aspergillus fumigatus*.** *Fungal Genet Biol* 2005, **42**(7):638-645.
72. Zhang X, Feschotte C, Zhang Q, Jiang N, Eggleston WB, Wessler SR: **P instability factor: an active maize transposon system associated with the amplification of Tourist-like MITEs and a new superfamily of transposases.** *Proc Natl Acad Sci USA* 2001, **98**(22):12572-12577.
73. Robert-Seilaniantz A, Navarro L, Bari R, Jones JD: **Pathological hormone imbalances.** *Curr Opin Plant Biol* 2007, **10**(4):372-379.
74. Sun S-K, Snyder WC: **The bakanae disease of the rice plant.** In *Fusarium: diseases, biology and taxonomy* Edited by: Nelson TAT PE, Cook RJ. The Pennsylvania State University Press; 1981:104-113.
75. Rademacher W, Graebe JE: **Gibberellin A4 produced by *Sphaceloma manihoticola*, the cause of the superelongation disease of cassava (*Manihot esculenta*).** *Biochem Biophys Res Commun* 1979, **91**(1):35-40.
76. Tudzynski B, Kawaide H, Kamiya Y: **Gibberellin biosynthesis in *Gibberella fujikuroi*: cloning and characterization of the copalyl diphosphate synthase gene.** *Curr Genet* 1998, **34**(3):234-240.
77. Bastos CN, Andebrhan T: **Presença de giberelina em basidiósporos de *Crinipellis perniciosa* (Stahel) Singer.** *Fitopatologia Brasileira* 1981, **6**:417-423.
78. Kilaru A, Bailey BA, Hasenstein KH: ***Moniliophthora perniciosa* produces hormones and alters endogenous auxin and salicylic acid in infected cocoa leaves.** *FEMS Microbiol Lett* 2007, **274**(2):238-244.
79. Wolf FT: **The Production of Indole Acetic Acid by *Ustilago Zeae*, and Its Possible Significance in Tumor Formation.** *Proc Natl Acad Sci USA* 1952, **38**(2):106-111.
80. Maor R, Haskin S, Levi-Kedimi H, Sharon A: **In planta production of indole-3-acetic acid by *Colletotrichum gloeosporioides* f. sp. aeschynomene.** *Appl Environ Microbiol* 2004, **70**(3):1852-1854.
81. Perley JE, Stowe BB: **On the ability of *Taphrina deformans* to produce indoleacetic acid from tryptophan by way of tryptamine.** *Plant Physiol* 1966, **41**(2):234-237.
82. Prusty R, Grisafi P, Fink GR: **The plant hormone indoleacetic acid induces invasive growth in *Saccharomyces cerevisiae*.** *Proc Natl Acad Sci USA* 2004, **101**(12):4153-4157.
83. Ross JJ, O'Neill DP, Smith JJ, Kerckhoffs LH, Elliott RC: **Evidence that auxin promotes gibberellin A1 biosynthesis in pea.** *Plant J* 2000, **21**(6):547-552.
84. Pierik R, Tholen D, Poorter H, Visser EJ, Voesenek LA: **The Janus face of ethylene: growth inhibition and stimulation.** *Trends Plant Sci* 2006, **11**(4):176-183.
85. Swarup R, Parry G, Graham N, Allen T, Bennett M: **Auxin cross-talk: integration of signalling pathways to control plant development.** *Plant Mol Biol* 2002, **49**(3-4):411-426.
86. Steffens B, Sauter M: **Epidermal cell death in rice is regulated by ethylene, gibberellin, and abscisic acid.** *Plant Physiol* 2005, **139**(2):713-721.
87. Pierik R, Cuppens ML, Voesenek LA, Visser EJ: **Interactions between ethylene and gibberellins in phytochrome-mediated shade avoidance responses in tobacco.** *Plant Physiol* 2004, **136**(2):2928-2936.
88. Maor R, Shirasu K: **The arms race continues: battle strategies between plants and fungal pathogens.** *Curr Opin Microbiol* 2005, **8**(4):399-404.
89. Dodds PN, Lawrence GJ, Catanzariti AM, Ayliffe MA, Ellis JG: **The *Melampora lini* AvrL567 avirulence genes are expressed in haustoria and their products are recognized inside plant cells.** *Plant Cell* 2004, **16**(3):755-768.
90. Kemen E, Kemen AC, Rafiqi M, Hempel U, Mendgen K, Hahn M, Voegelé RT: **Identification of a protein from rust fungi transferred from haustoria into infected plant cells.** *Mol Plant Microbe Interact* 2005, **18**(11):1130-1139.
91. Rooney HC, Van't Klooster JW, Hoorn RA van der, Joosten MH, Jones JD, de Wit PJ: ***Cladosporium Avr2* inhibits tomato Rcr3 protease required for Cf-2-dependent disease resistance.** *Science* 2005, **308**(5729):1783-1786.
92. Kamper J, Kahmann R, Bolker M, Ma LJ, Brefort T, Saville BJ, Banuett F, Kronstad JW, Gold SE, Muller O, Perlin MH, Wosten HA, de Vries R, Ruiz-Herrera J, Reynaga-Pena CG, Snetselaar K, McCann M, Perez-Martin J, Feldbrugge M, Basse CW, Steinberg G, Ibeas JI, Holloman W, Guzman P, Farman M, Stajich JE, Sentandreu R, Gonzalez-Prieto JM, Kennell JC, Molina L, Schirawski J, Mendoza-Mendoza A, Greilinger D, Munch K, Rossel N, Scherer M, Vranes M, Ladendorff O, Vincon V, Fuchs U, Sandrock B, Meng S, Ho EC, Cahill MJ, Boyce KJ, Klose J, Klosterman SJ, Deelstra HJ, Ortiz-Castellanos L, Li W, Sanchez-Alonso P, Schreier PH, Hauser-Hahn I, Vaupel M, Koopmann E, Friedrich G, Voss H, Schluter T, Margolis J, Platt D, Swimmer C, Gnirke A, Chen F, Vysotskaia V, Mannhaupt G, Guldener U, Munsterkotter M, Haase D, Oosterheld M, Mewes HW, Mauceli EW, DeCaprio D, Wade CM, Butler J, Young S, Jaffe DB, Calvo S, Nusbaum C, Galagan J, Birren BW: **Insights from the genome of the biotrophic fungal plant pathogen *Ustilago maydis*.** *Nature* 2006, **444**(7115):97-101.
93. Rep M: **Small proteins of plant-pathogenic fungi secreted during host colonization.** *FEMS Microbiol Lett* 2005, **253**(1):19-27.
94. Zapparoli G, Garcia O, Medrano FJ, Tiburcio R, Costa GL, Pereira GAG: **Identification of a second family of genes in *Moniliophthora perniciosa*, the causal agent of Witches' Broom disease in cacao, encoding necrosis inducing proteins similar to cerato platanins.** *Mycol Res* 2008.
95. Nonaka T, Dohmae N, Hashimoto Y, Takio K: **Amino acid sequences of metalloendopeptidases specific for acyl-lysine bonds from *Grifola frondosa* and *Pleurotus ostreatus* fruiting bodies.** *J Biol Chem* 1997, **272**(48):30032-30039.
96. Sepcic K, Berne S, Potrich C, Turk T, Macek P, Menestrina G: **Interaction of ostreolysin, a cytolytic protein from the edible mushroom *Pleurotus ostreatus*, with lipid membranes and modulation by lysophospholipids.** *Eur J Biochem* 2003, **270**(6):1199-1210.

97. Wymelenberg A, Vanden, Minges P, Sabat G, Martinez D, Rajangam AS, Teeri TT, Gaskell J, Kersten PJ, Cullen D: **The *Phanerochaete chrysosporium* secretome: database predictions and initial mass spectrometry peptide identifications in cellulose-grown medium.** *J Biotechnol* 2005, **118**(1):17-34.
98. Garcia-Conesa MT, Kroon PA, Ralph J, Mellon FA, Colquhoun IJ, Saulnier L, Thibault JF, Williamson G: **A cinnamoyl esterase from *Aspergillus niger* can break plant cell wall cross-links without release of free diferulic acids.** *Eur J Biochem* 1999, **266**(2):644-652.
99. van Loon LC, Rep M, Pieterse CM: **Significance of inducible defense-related proteins in infected plants.** *Annu Rev Phytopathol* 2006, **44**:135-162.
100. Broekaert WF, Terras FRG, BPA C: **Induced and preformed antimicrobial proteins.** In *Mechanisms of Resistance to Plant Diseases* Edited by: Slusarenko AJFR, Van Loon LC. Dordrecht: Kluwer; 2000:371-477.
101. Hu X, Reddy AS: **Nucleotide sequence of a cDNA clone encoding a thaumatin-like protein from *Arabidopsis*.** *Plant Physiol* 1995, **107**(1):305-306.
102. Punja Z: **Transgenic carrots expressing a thaumatin-like protein display enhanced resistance to several fungal pathogens.** *Can J Plant Pathol* 2005, **27**:291-296.
103. Fierens E, Rombouts S, Gebruers K, Goesaert H, Brijis K, Beaugrand J, Volckaert G, van Campenhout S, Proost P, Courtin CM, Delcour JA: **TLXI, a novel type of xylanase inhibitor from wheat (*Triticum aestivum*) belonging to the thaumatin family.** *Biochem J* 2007, **403**(3):583-591.
104. Miosga T, Schaaff-Gerstenschlager I, Chalwatzis N, Baur A, Boles E, Fournier C, Schmitt S, Velten C, Wilhelm N, Zimmermann FK: **Sequence analysis of a 33.1 kb fragment from the left arm of *Saccharomyces cerevisiae* chromosome X, including putative proteins with leucine zippers, a fungal Zn(II)-Cys6 binuclear cluster domain and a putative alpha 2-SCB-alpha 2 binding site.** *Yeast* 1995, **11**(7):681-689.
105. Sakamoto Y, Watanabe H, Nagai M, Nakade K, Takahashi M, Sato T: ***Lentinula edodes* tlgI encodes a thaumatin-like protein that is involved in lentinan degradation and fruiting body senescence.** *Plant Physiol* 2006, **141**(2):793-801.
106. Chaves FC, Gianfagna TJ: **Necrotrophic phase of *Monilophthora perniciosa* causes salicylic acid accumulation in infected stems of cacao.** *Physiol Mol Plant Pathol* 2006, **69**(1-3):104-108.
107. Aro N, Pakula T, Penttila M: **Transcriptional regulation of plant cell wall degradation by filamentous fungi.** *FEMS Microbiol Rev* 2005, **29**(4):719-739.
108. Cuomo CA, Guidener U, Xu JR, Trail F, Turgeon BG, Di Pietro A, Walton JD, Ma LJ, Baker SE, Rep M, Adam G, Antoniw J, Baldwin T, Calvo S, Chang YL, Decaprio D, Gale LR, Gnerre S, Goswami RS, Hammond-Kosack K, Harris LJ, Hilburn K, Kennell JC, Kroken S, Magnuson JK, Mannhaupt G, Mauceli E, Mewes HW, Mitterbauer R, Muehlbauer G, Munsterkotter M, Nelson D, O'Donnell K, Ouellet T, Qi W, Quesneville H, Roncero MI, Seong KY, Tetko IV, Urban M, Waalwijk C, Ward TJ, Yao J, Birren BW, Kistler HC: **The *Fusarium graminearum* genome reveals a link between localized polymorphism and pathogen specialization.** *Science* 2007, **317**(5843):1400-1402.
109. Millar AH, Day DA: **Nitric oxide inhibits the cytochrome oxidase but not the alternative oxidase of plant mitochondria.** *FEBS Lett* 1996, **398**(2-3):155-158.
110. Juarez O, Guerra G, Martinez F, Pardo JP: **The mitochondrial respiratory chain of *Ustilago maydis*.** *Biochim Biophys Acta* 2004, **1658**(3):244-251.
111. Juarez O, Guerra G, Velazquez I, Flores-Herrera O, Rivera-Perez RE, Pardo JP: **The physiologic role of alternative oxidase in *Ustilago maydis*.** *FEBS J* 2006, **273**(20):4603-4615.
112. Joseph-Horne T, Hollomon DW, Wood PM: **Fungal respiration: a fusion of standard and alternative components.** *Biochim Biophys Acta* 2001, **1504**(2-3):179-195.
113. Idnurm A, Howlett BJ: **Isocitrate lyase is essential for pathogenicity of the fungus *Leptosphaeria maculans* to canola (*Brassica napus*).** *Eukaryot Cell* 2002, **1**(5):719-724.
114. Sakai S, Nishide T, Munir E, Baba K, Inui H, Nakano Y, Hattori T, Shimada M: **Subcellular localization of glyoxylate cycle key enzymes involved in oxalate biosynthesis of wood-destroying basidiomycete *Fomitopsis palustris* grown on glucose.** *Microbiology* 2006, **152**(Pt 6):1857-1866.
115. Wang ZY, Thornton CR, Kershaw MJ, Debaio L, Talbot NJ: **The glyoxylate cycle is required for temporal regulation of virulence by the plant pathogenic fungus *Magnaporthe grisea*.** *Mol Microbiol* 2003, **47**(6):1601-1612.
116. Segers G, Bradshaw N, Archer D, Blissett K, Oliver RP: **Alcohol oxidase is a novel pathogenicity factor for *Cladosporium fulvum*, but aldehyde dehydrogenase is dispensable.** *Mol Plant Microbe Interact* 2001, **14**(3):367-377.
117. Daniel G, Volc J, Filonova L, Pilhal O, Kubatova E, Halada P: **Characteristics of *Gloeophyllum trabeum* Alcohol Oxidase, an Extracellular Source of H₂O₂ in Brown Rot Decay of Wood.** *Appl Environ Microbiol* 2007, **73**(19):6241-6253.
118. Nakagawa T, Miyaji T, Yurimoto H, Sakai Y, Kato N, Tomizuka N: **A methylotrophic pathway participates in pectin utilization by *Candida boidinii*.** *Appl Environ Microbiol* 2000, **66**(10):4253-4257.
119. Fall R, Benson AA: **Leaf methanol - The simplest natural product from plants.** *Trends Plant Sci* 1996, **1**(9):296-301.
120. Abanda-Nkpaw D, Musch M, Tschiersch J, Boettner M, Schwab W: **Molecular interaction between *Methylobacterium extorquens* and seedlings: growth promotion, methanol consumption, and localization of the methanol emission site.** *J Exp Bot* 2006, **57**(15):4025-4032.
121. Wei Y, Shen W, Dauk M, Wang F, Selvaraj G, Zou J: **Targeted gene disruption of glycerol-3-phosphate dehydrogenase in *Colletotrichum gloeosporioides* reveals evidence that glycerol is a significant transferred nutrient from host plant to fungal pathogen.** *J Biol Chem* 2004, **279**(1):429-435.
122. Santos RX, Melo SC, Cascardo JC, Brendel M, Pungartnik C: **Carbon source-dependent variation of acquired mutagen resistance of *Monilophthora perniciosa*: Similarities in natural and artificial systems.** *Fungal Genet Biol* 2008.
123. Nakayama A, Park S, Zheng-Jun X, Nakajima M, Yamaguchi I: **Immunohistochemistry of active gibberellins and gibberellin-inducible alpha-amylase in developing seeds of morning glory.** *Plant Physiol* 2002, **129**(3):1045-1053.
124. Pellier AL, Lauge R, Veneault-Fourrey C, Langin T: **CLNR1, the AREA/NIT2-like global nitrogen regulator of the plant fungal pathogen *Colletotrichum lindemuthianum* is required for the infection cycle.** *Mol Microbiol* 2003, **48**(3):639-655.
125. Divon HH, Fluhr R: **Nutrition acquisition strategies during fungal infection of plants.** *FEMS Microbiol Lett* 2007, **266**(1):65-74.
126. Marzluf GA: **Genetic regulation of nitrogen metabolism in the fungi.** *Microbiol Mol Biol Rev* 1997, **61**(1):17-32.
127. Lau G, Hamer JE: **Regulatory genes controlling MPG1 expression and pathogenicity in the rice blast fungus *Magnaporthe grisea*.** *Plant Cell* 1996, **8**(5):771-781.
128. Ewing B, Hillier L, Wendt MC, Green P: **Base-calling of automated sequencer traces using phred. I. Accuracy assessment.** *Genome Res* 1998, **8**(3):175-185.
129. Chou HH, Holmes MH: **DNA sequence quality trimming and vector removal.** *Bioinformatics* 2001, **17**(12):1093-1104.
130. Usuka J, Zhu W, Brendel V: **Optimal spliced alignment of homologous cDNA to a genomic DNA template.** *Bioinformatics* 2000, **16**(3):203-211.
131. Gremme G, Brendel V, Sparks ME, Kurtz S: **Engineering a software tool for gene structure prediction in higher organisms.** *Inform Software Technol* 2005, **47**(15):965-978.
132. Lowe TM, Eddy SR: **tRNAscan-SE: A program for improved detection of transfer RNA genes in genomic sequence.** *Nucleic Acids Res* 1997, **25**(5):955-964.
133. Karp PD, Paley S, Romero P: **The Pathway Tools software.** *Bioinformatics* 2002, **18**(Suppl 1):S225-232.
134. Nielsen H, Engelbrecht J, Brunak S, von Heijne G: **Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites.** *Protein Eng* 1997, **10**(1):1-6.
135. Collart MA, Oliviero S: **Preparation of yeast RNA by extraction with hot acidic phenol.** In *Current Protocols in Molecular Biology* Edited by: Ausubel FM, Brent R, Kingston RE, Moore DD, Seidman JG, Smith JA, Hoboken KS. New Jersey: John Wiley & Sons, Inc; 2006.

Protein

GenPept

hypothetical protein MPER_07801 [Moniliophthora perniciosa FA553]

GenBank: EEB93525.1

[Identical Proteins](#) [FASTA](#) [Graphics](#)

Go to:

LOCUS EEB93525 343 aa linear PLN 02-DEC-2008
 DEFINITION hypothetical protein MPER_07801 [Moniliophthora perniciosa FA553].
 ACCESSION EEB93525
 VERSION EEB93525.1
 DBLINK BioProject: [PRJNA28951](#)
 BioSample: [SAMN02953729](#)
 DBSOURCE accession [ABRE01012504.1](#)
 KEYWORDS
 SOURCE Moniliophthora perniciosa FA553
 ORGANISM [Moniliophthora perniciosa FA553](#)
 Eukaryota; Fungi; Dikarya; Basidiomycota; Agaricomycotina;
 Agaricomycetes; Agaricomycetidae; Agaricales; Marasmiaceae;
 Moniliophthora.
 REFERENCE 1 (residues 1 to 343)
 AUTHORS Mondego,J.M., Carazzolle,M.F., Costa,G.G., Formighieri,E.F.,
 Parizzi,L.P., Rincones,J., Cotomacci,C., Carraro,D.M., Cunha,A.F.,
 Carrer,H., Vidal,R.O., Estrela,R.C., Garcia,O., Thomazella,D.P., de
 Oliveira,B.V., Pires,A.B., Rio,M.C., Araujo,M.R., de Moraes,M.H.,
 Castro,L.A., Gramacho,K.P., Goncalves,M.S., Neto,J.P., Neto,A.G.,
 Barbosa,L.V., Gultinan,M.J., Bailey,B.A., Meinhardt,L.W.,
 Cascardo,J.C. and Pereira,G.A.
 TITLE A genome survey of Moniliophthora perniciosa gives new insights
 into Witches' Broom Disease of cacao
 JOURNAL BMC Genomics 9, 548 (2008)
 PUBMED [19019209](#)
 REMARK Publication Status: Online-Only
 REFERENCE 2 (residues 1 to 343)
 AUTHORS Mondego,J.M.C., Carazzolle,M.F., Costa,G.G.L., Formighieri,E.F.,
 Parizzi,L.P., Rincones,J., Cotomacci,C., Carraro,D.M., Cunha,A.F.,
 Carrer,H., Vidal,R.O., Estrela,R.C., Garcia,O., Thomazella,D.P.T.,
 de Oliveira,B.V., Pires,A.B.L., Rio,M.C.S., Araujo,M.R.R., de
 Moraes,M.H., Castro,L.A.B., Gramacho,K.P., Goncalves,M.S.,
 Neto,A.G., Barbosa,L.V., Gultinan,M.J., Bailey,B.A.,
 Meinhardt,L.W., Cascardo,J.C.M. and Pereira,G.A.G.
 CONSRMT Moniliophthora perniciosa genome project
 TITLE Direct Submission
 JOURNAL Submitted (19-AUG-2008) Laboratorio de Genomica e Expressao,
 Departamento de Genetica e Evolucao, Instituto de Biologia,
 Universidade Estadual de Campinas, Cidade Universitaria Zeferino
 Vaz, Campinas, Sao Paulo 13083-970, Brasil
 COMMENT Contigs ABRE01001187 ABRE01003672 ABRE01009875 ABRE01010381
 ABRE01011006 ABRE01011419 ABRE01012265 ABRE01018553 ABRE01018752
 ABRE01022003 ABRE01024029 ABRE01024864 were updated to correct the
 translations of their CDS in November 2015.
 ##Genome-Assembly-Data-START##
 Assembly Method :: PHRED-PHRAP v. 2007
 Genome Coverage :: 1.9x
 Sequencing Technology :: ABI 3700
 ##Genome-Assembly-Data-END##
 Method: conceptual translation.
 FEATURES
 source 1..343
 /organism="Moniliophthora perniciosa FA553"
 /strain="FA553"
 /db_xref="taxon:554373"
 /note="biotype: C"
 Protein 1..343
 /product="hypothetical protein"
 Region 5..>268
 /region_name="Atrophin-1"
 /note="Atrophin-1 family; cl26464"
 /db_xref="CDD:331285"
 CDS 1..343
 /locus_tag="MPER_07801"
 /coded_by="complement(join(ABRE01012504.1:287..392,

. 9/27/2017

hypothetical protein MPER_07801 [Monilophthora pernicioso FA553] - Protein - NCBI

ABRE01012504.1:410..1146,ABRE01012504.1:1233..1286,
ABRE01012504.1:1345..1479))"
/note="AutoFACTv3.4 (automatic product
annotation):Unassigned protein"

ORIGIN

1 mkklmgngrn ssdrdtasin ghkksdsiis dqtktwsssf tqtdssslq efddlmrsda
61 tmkvykqekq eknqrqndnr hptssqplri dnrrpslvrn vdsiqedeee sqikssssaa
121 vsqrvrqisv tpskpsian rarsisaagq sipsligrks sktgiasppl ptmqtpdkra
181 vgmppmgfggd pfpqktrkiq hnresldldd vmagsdedkm avavpkpstp nkrpivkpag
241 gvssstreim dflnegppem pgrplgrpgr efpefmdngs idiglndqtg skskagrlqr
301 mmsklisiggp erpdagrlfk edvytyitid fiskwnqpen lcs

//

(12) STANDARD PATENT
(19) AUSTRALIAN PATENT OFFICE

(11) Application No. **AU 2017261812 B2**

(54) Title
Adeno-associated virus variant capsids and methods of use thereof

(51) International Patent Classification(s)
C07K 14/075 (2006.01) **C12N 15/63** (2006.01)
A61K 48/00 (2006.01) **C12N 15/861** (2006.01)
C12N 15/62 (2006.01) **C12Q 1/70** (2006.01)

(21) Application No: **2017261812** (22) Date of Filing: **2017.05.12**

(87) WIPO No: **WO17/197355**

(30) Priority Data

(31) Number	(32) Date	(33) Country
62/384,590	2016.09.07	US
62/454,612	2017.02.03	US
62/336,441	2016.05.13	US
62/378,106	2016.08.22	US

(43) Publication Date: **2017.11.16**

(44) Accepted Journal Date: **2019.04.18**

(71) Applicant(s)
4D Molecular Therapeutics Inc.

(72) Inventor(s)
Kirn, David H.;Kotterman, Melissa;Schaffer, David

(74) Agent / Attorney
WRAYS PTY LTD, L7 863 Hay St, Perth, WA, 6000, AU

(56) Related Art
WO 2018/022905 A2
WO 2017/218842 A1

WHAT IS CLAIMED IS:

1. A variant adeno-associated virus (AAV) capsid protein comprising a peptide insertion in the GH-loop of the capsid protein relative to a corresponding parental AAV capsid protein, wherein the peptide insertion comprises the amino acid sequence ISDQTKH (SEQ ID NO:14).
2. The variant AAV capsid protein of claim 1, wherein the peptide insertion comprises the amino acid sequence $Y_1Y_2ISDQTKHY_3$, wherein each of Y_1 - Y_3 is independently selected from Ala, Leu, Gly, Ser, Thr, and Pro.
3. The variant AAV capsid protein of claim 2, wherein the peptide insertion comprises or consists of the amino acid sequence LAISDQTKHA (SEQ ID NO:28).
4. The variant AAV capsid protein of any one of claims 1-3, wherein the insertion site is located between amino acids corresponding to amino acids 587 and 588 of VP1 of AAV2 (SEQ ID NO:2) or the corresponding position in the capsid protein of another AAV serotype.
5. The variant AAV capsid protein of any one of claims 1-4, wherein the capsid protein comprises one or more amino acid substitutions relative to VP1 of AAV2 (SEQ ID NO:2) or one or more corresponding substitutions in the capsid protein of another AAV serotype.
6. The variant AAV capsid protein of any one of claims 1-5, wherein the capsid protein comprises one or more of the following amino acid substitutions relative to VP1 of AAV2 (SEQ ID NO:2) or the corresponding substitution(s) in the capsid protein of another AAV serotype: M1L, L15P, P34A, N57D, N66K, R81Q, Q101R, S109T, R144K, R144M, Q164K, T176P, L188I, S196Y, G226E, G236V, I240T, P250S, N312K, P363L, D368H, N449D, T456K, S463Y, D472N, R484C, A524T, P535S, N551S, A593E, I698V, V708I, V719M, S721L, and L735Q.
7. The variant AAV capsid protein of claim 6, wherein the capsid protein comprises a P34A amino acid substitution relative to VP1 of AAV2 (SEQ ID NO:2) or the corresponding substitution in the capsid protein of another AAV serotype.

8. The variant AAV capsid protein of claim 7, wherein the capsid protein comprises an amino acid sequence at least 90% identical, at least 95% identical, at least 99% identical or 100% identical to the entire length of the amino acid sequence set forth as SEQ ID NO:42.

9. The variant AAV capsid protein of claim 8, wherein the capsid protein consists essentially of the amino acid sequence set forth as SEQ ID NO:42.

10. The variant AAV capsid according to any one of claims 1-9, wherein the capsid protein confers to an infectious recombinant AAV (rAAV) virion an increased infectivity of a retinal cell compared to the infectivity of the retinal cell by an AAV virion comprising the corresponding parental AAV capsid protein.

11. An isolated nucleic acid comprising a nucleotide sequence that encodes a variant AAV capsid protein according to any one of claims 1-10.

12. An infectious rAAV virion comprising a variant AAV capsid protein according to any one of claims 1-10.

13. The rAAV virion according to claim 12, further comprising a heterologous nucleic acid encoding a gene product.

14. The rAAV virion according to claim 13, wherein the gene product is a protein or short interfering RNA.

15. The rAAV virion according to claim 14, wherein the variant AAV capsid protein comprises an amino acid sequence having at least 90%, at least 95%, or at least 99% amino acid sequence identity to the amino acid sequence set forth in SEQ ID NO:42 and wherein the heterologous nucleic acid comprises a nucleotide sequence encoding a Rab escort protein-1 protein.

16. The rAAV virion according to claim 15, wherein the variant AAV capsid consists of the amino acid sequence set forth as SEQ ID NO:42 and wherein the nucleotide sequence encoding the Rab escort protein-1 protein is operably linked to a CAG promoter.

17. The rAAV virion according to claim 14, wherein the variant AAV capsid protein comprises an amino acid sequence having at least 90%, at least 95%, or at least 99% amino acid sequence identity to the amino acid sequence set forth in SEQ ID NO:42 and wherein the

heterologous nucleic acid comprises a nucleotide sequence encoding a retinitis pigmentosa GTPase regulator protein.

18. The rAAV virion according to claim 17, wherein the variant AAV capsid protein comprises or consists of the amino acid sequence set forth as SEQ ID NO:42.

19. The rAAV virion according to claim 18, wherein the nucleotide sequence encoding the retinitis pigmentosa GTPase regulator protein is operably linked to a rhodopsin kinase promoter.

20. The rAAV virion according to claim 14, wherein the variant AAV capsid protein comprises an amino acid sequence having at least 90%, at least 95%, or at least 99% amino acid identity to the amino acid sequence set forth in SEQ ID NO:42 and wherein the heterologous nucleic acid comprises a nucleotide sequence encoding a polypeptide that inhibits the activity of vascular endothelial growth factor (VEGF).

21. The rAAV virion according to claim 20, wherein the variant AAV capsid protein comprises or consists of the amino acid sequence set forth as SEQ ID NO:42.

22. The rAAV virion according to claim 21, wherein the polypeptide is a fusion protein or antibody.

23. A pharmaceutical composition comprising an rAAV virion according to any one of claims 14 to 22 and a pharmaceutically acceptable excipient.

24. A method for delivering a heterologous nucleic acid to a retinal cell comprising contacting the retinal cell with an rAAV virion according to any one of claims 14 to 22 or a pharmaceutical composition according to claim 23.

25. The method according to claim 24, wherein the retinal cell is selected from the group consisting of retinal ganglion cells, amacrine cells, horizontal cells, bipolar cells, photoreceptor cells, Müller glial cells, microglial cells and retinal pigment epithelium.

26. The use of an rAAV virion according to any one of claims 14 to 22 or a pharmaceutical composition according to claim 23 for the manufacture of a medicament.

27. The use of an rAAV virion according to claim 15 or 16 or a pharmaceutical composition comprising an rAAV virion according to claim 15 or 16, for the manufacture of a medicament for the treatment of choroideremia.

28. A method to treat choroideremia in a subject in need thereof comprising administering to the subject an effective amount of an rAAV virion according to claim 15 or 16 or a pharmaceutical composition comprising an rAAV virion according to claim 15 or 16.

29. The use of an rAAV virion according to any one of claims 17 to 19 or a pharmaceutical composition comprising an rAAV virion according to any one of claims 17 to 19 for the manufacture of a medicament for the treatment of X-linked retinitis pigmentosa.

30. A method to treat X-linked retinitis pigmentosa in a subject in need thereof comprising administering to the subject an effective amount of an rAAV virion according to any one of claims 17 to 19 or a pharmaceutical composition comprising an rAAV virion according to any one of claims 17 to 19.

31. The use of an rAAV virion according to any one of claims 20 to 22 or a pharmaceutical composition comprising an rAAV virion according to any one of claims 20 to 22, for the manufacture of a medicament for the treatment of wet age-related macular degeneration, diabetic retinopathy, diabetic macular edema, or choroidal neovascularization.

32. A method to treat wet age-related macular degeneration, diabetic retinopathy, diabetic macular edema, or choroidal neovascularization in a subject in need thereof, comprising administering to the subject an effective amount of an rAAV virion according to any one of claims 20 to 22 or a pharmaceutical composition comprising an rAAV virion according to any one of claims 20 to 22.

33. The use of an rAAV virion or pharmaceutical composition according to any one of claims 26, 27, 29, and 31, wherein the rAAV virion or pharmaceutical composition is for administration by intravitreal injection.

34. The method according to any one of claims 28, 30 and 32, wherein the rAAV virion or pharmaceutical composition is administered to the subject by intravitreal injection.

35. The use of an rAAV virion or pharmaceutical composition according to any one of claims 26, 27, 29 and 31 wherein the medicament is for treating a human subject.

36. The method according to any one of claims 28, 30 and 32, wherein the subject is a human.

REIVINDICACIONES

1. Una proteína de la cápside del virus adenoasociado (AAV) variante que comprende una inserción peptídica en el bucle GH de la proteína de la cápside con respecto a una proteína de la cápside del AAV parental correspondiente, en donde la inserción peptídica comprende la secuencia de aminoácidos ISDQTKH (SEQ ID NO:14).

2. La proteína de la cápside del AAV variante de acuerdo con la reivindicación 1, en donde la inserción peptídica comprende la secuencia de aminoácidos $Y_1Y_2ISDQTKHY_3$, en donde cada Y_1 - Y_3 se selecciona de manera independiente de Ala, Leu, Gly, Ser, Thr, y Pro.

3. La proteína de la cápside del AAV variante de acuerdo con la reivindicación 2, en donde la inserción peptídica comprende o consiste en la secuencia de aminoácidos LAISDQTKHA (SEQ ID NO:28).

4. La proteína de la cápside del AAV variante de cualquiera de las reivindicaciones 1-3, en donde el sitio de inserción se localiza entre los aminoácidos correspondientes a los aminoácidos 587 y 588 en VP1 del AAV2 (SEQ ID NO:2) o la posición correspondiente en la proteína de la cápside de otro serotipo de AAV.

5. La proteína de la cápside del AAV variante de acuerdo con cualquiera de las reivindicaciones 1-4, en donde la proteína de la cápside comprende una o más sustituciones de aminoácidos con respecto a la VP1 del AAV2 (SEQ ID NO:2) o una o más sustituciones correspondientes en la proteína de la cápside de otro serotipo de AAV.

6. La proteína de la cápside del AAV variante de acuerdo con cualquiera de las reivindicaciones 1-5, en donde la proteína de la cápside comprende una o más de las siguientes sustituciones de aminoácidos con respecto a la VP1 del AAV2 (SEQ ID NO:2) o la(s) sustitución(es) correspondiente(s) a la proteína de la cápside de otro serotipo de AAV: M1L, L15P, P34A, N57D, N66K, R81Q, Q101R, S109T, R144K, R144M, Q164K, T176P, L188I, S196Y, G226E, G236V, I240T, P250S, N312K, P363L, D368H, N449D, T456K, S463Y, D472N, R484C, A524T, P535S, N551S, A593E, I698V, V708I, V719M, S721L, y L735Q..

7. La proteína de la cápside del AAV variante de acuerdo con la reivindicación 6, en donde la proteína de la cápside comprende una sustitución de aminoácido P34A con respecto a la VP1 del AAV2 (SEQ ID NO:2) o la sustitución correspondiente en la proteína de la cápside de otro serotipo de AAV.

8. La proteína de la cápside del AAV variante de acuerdo con la reivindicación 7, en donde la proteína de la cápside comprende una secuencia de aminoácidos al menos

90 % idéntica, al menos 95 % idéntica, al menos 99 % idéntica o 100% idéntica a la longitud completa de la secuencia de aminoácidos descrita en SEQ ID NO:42.

9. La proteína de la cápside del AAV variante de acuerdo con la reivindicación 8, en donde la proteína de la cápside consiste esencialmente de una secuencia de aminoácidos descrita en SEQ ID NO:42.

10. La cápside del AAV variante de acuerdo con cualquiera de las reivindicaciones 1-9, en donde la proteína de la cápside confiere a un virión AAV recombinante infeccioso (rAAV) una infectividad aumentada de una célula retiniana comparada con la infectividad de la célula retiniana por un virión AAV que comprende la proteína de la cápside del AAV parental correspondiente.

11. Un ácido nucleído aislado que comprende una secuencia de nucleótidos que codifica una proteína de la cápside del AAV variante de acuerdo con cualquiera de las reivindicaciones 1-10.

12. Un virión infeccioso (rAAV) que comprende una proteína de la cápside del AAV variante de acuerdo con cualquiera de las reivindicaciones 1-10.

13. El virión rAAV de acuerdo con la reivindicación 12, que además comprende un ácido nucleico heterólogo que codifica un producto génico.

14. El virión rAAV de acuerdo con la reivindicación 13, en donde el producto génico es una proteína o un ARN de interferencia corto.

15. El virión rAAV de acuerdo con la reivindicación 14, en donde la proteína de la cápside del AAV variante comprende una secuencia de aminoácidos que posee al menos 90%, al menos 95%, o al menos 99% identidad de aminoácidos con la secuencia de aminoácidos descrita en la SEQ ID NO:42 y en donde el ácido nucleico heterólogo comprende una secuencia de nucleótidos que codifica una proteína escolta Rab-1.

16. El virión rAAV de acuerdo con la reivindicación 15, en donde la cápside del AAV variante consiste en la secuencia de aminoácidos descrita en SEQ ID NO: 42 y en donde la secuencia de nucleótidos que codifica la proteína escolta Rab-1 está unida operativamente a un promotor CAG.

17. El virión rAAV de acuerdo con la reivindicación 14, en donde la proteína de la cápside de AAV variante comprende una secuencia de aminoácidos que posee al menos 90%, al menos 95%, o al menos 99% de identidad de aminoácidos con la secuencia de aminoácidos descrita en la SEQ ID NO:42 y en donde el ácido nucleico heterólogo comprende una secuencia de nucleótidos que codifica una proteína reguladora de la GTPasa de retinitis pigmentosa.

18. El virión rAAV de acuerdo con la reivindicación 17, en donde la proteína de cápside del AAV variante comprende o consiste en la secuencia de aminoácidos que se describe en SEQ ID NO: 42.

19. El virión rAAV de acuerdo con la reivindicación 18, en donde la secuencia de nucleótidos que codifica la proteína reguladora de la GTPasa de retinitis pigmentosa está operativamente unida con un promotor de rodopsina quinasa.

20. El virión rAAV de acuerdo con la reivindicación 14, en donde la proteína de la cápside del AAV variante comprende una secuencia de aminoácidos que posee al menos un 90%, al menos un 95%, o al menos un 99% de identidad de aminoácidos con la secuencia de aminoácidos que se describe en la SEQ ID NO:42 y en donde el ácido nucleico heterólogo comprende una secuencia de nucleótidos que codifica un polipéptido que inhibe la actividad del factor de crecimiento endotelial vascular (VEGF).

21. El virión rAAV de acuerdo con la reivindicación 20, en donde la proteína de cápside del AAV variante comprende o consiste en la secuencia de aminoácidos que se describe en SEQ ID NO: 42.

22. El virión rAAV de acuerdo con la reivindicación 21, en donde el polipéptido es una proteína de fusión o un anticuerpo.

23. Una composición farmacéutica que comprende un virión rAAV de acuerdo con cualquiera de las reivindicaciones 14-22 y un excipiente farmacéuticamente aceptable.

24. Un método para administrar un ácido nucleico heterólogo a una célula retiniana que comprende contactar la célula retiniana con un virión rAAV de acuerdo con cualquiera de las reivindicaciones 14-22 o una composición farmacéutica de acuerdo con la reivindicación 23.

25. El método de acuerdo con la reivindicación 24, en donde la célula retiniana, se selecciona del grupo que consiste en células ganglionares retinianas, células amacrinas, células horizontales, células bipolares, células fotorreceptoras, células gliales de Müller, células microgliales y epitelio pigmentar retinal.

26. El uso de un virión rAAV de acuerdo con cualquiera de las reivindicaciones 14 a 22 o una composición farmacéutica de acuerdo con la reivindicación 23 para la fabricación de un medicamento.

27. El uso de un virión rAAV de acuerdo con la reivindicación 15 o 16 o una composición farmacéutica que comprende un virión rAAV de acuerdo con la reivindicación 15 o 16, para la fabricación de un medicamento para el tratamiento de coroideremia.

28. Un método para tratar coroideremia en un sujeto que lo necesita que

comprende administrar al sujeto una cantidad efectiva de un virión rAAV de acuerdo con la reivindicación 15 o 16 o una composición farmacéutica que comprende un virión rAAV de acuerdo con la reivindicación 15 o 16.

29. El uso de un virión rAAV de acuerdo con cualquiera de las reivindicaciones 17 a 19 o una composición farmacéutica que comprende un virión rAAV de acuerdo con cualquiera de las reivindicaciones 17 a 19 para la fabricación de un medicamento para el tratamiento de retinitis pigmentosa ligadas al X.

30. Un método para tratar la retinitis pigmentosa ligadas al X en un sujeto que lo necesita que comprende administrar al sujeto una cantidad efectiva de un virión rAAV de acuerdo con cualquiera de las reivindicaciones 17 a 19 o una composición farmacéutica que comprende un virión rAAV de acuerdo con cualquiera de las reivindicaciones 17 a 19.

31. El uso de un virión rAAV de acuerdo con cualquiera de las reivindicaciones 20 a 22 o de una composición farmacéutica que comprende un virión rAAV de acuerdo con cualquiera de las reivindicaciones 20 a 22, para la fabricación de un medicamento para el tratamiento de degeneración macular húmeda relacionada con la edad, retinopatía diabética, edema macular diabético o neovascularización coroidal.

32. Un método para tratar la degeneración macular húmeda relacionada con la edad, retinopatía diabética, edema macular diabético o neovascularización coroidal en un sujeto que lo necesita, que comprende administrar al sujeto una cantidad efectiva de un virión rAAV de acuerdo con cualquiera de las reivindicaciones 20 a 22 o de una composición farmacéutica que comprende un virión rAAV de acuerdo con cualquiera de las reivindicaciones 20 a 22.

33. El uso de un virión rAAV o composición farmacéutica de acuerdo con cualquiera de las reivindicaciones 26, 27, 29 y 31, en donde el virión rAAV o composición farmacéutica es para administración por inyección intravítrea.

34. El método de acuerdo con cualquiera de las reivindicaciones 28, 30 y 32, en donde el virión rAAV o la composición farmacéutica se administra al sujeto por inyección intravítrea.

35. El uso de un virión rAAV o composición farmacéutica de acuerdo con cualquiera de las reivindicaciones 26, 27, 29 y 31 en donde el medicamento es para tratar a un sujeto humano.

36. El método de acuerdo con cualquiera de las reivindicaciones 28, 30 y 32, en donde el sujeto es un humano.



9 April 2019



Delivering a world leading IP system

Phone: 1300 651 010
International: +61 2 6283 2999

www.ipaustralia.gov.au

ABN: 38 113 072 755

Notice of acceptance for patent application

WRAYS PTY LTD

L7
863 Hay St
Perth WA 6000
Australia

Your reference	274011
Application number	2017261812
Applicant name	4D Molecular Therapeutics Inc.

Dear WRAYS PTY LTD,

This patent application was accepted on 5 April 2019. The accepted specification incorporates the following amendments:

S104 amendments up to and including item number: 3

The total number of claims at acceptance has been reported as: 36

Notice of the acceptance will appear in the Australian Official Journal of Patents on 18 April 2019.

Please review the acceptance details attached to this letter to ensure that they are correct. If you wish to amend any details prior to the grant please do so within 3 months of the advertisement of acceptance.

An Invitation to Pay (ITP) the acceptance fee will be issued to you. This fee will include an additional component if the number of claims exceeds 20.

Details of this patent application can be viewed on AusPat, our Australian patent search database.

Yours sincerely,

IP Australia

Acceptance summary

Standard patent details

Patent number: 2017261812
Title: Adeno-associated virus variant capsids and methods of use thereof
Your reference: 274011
Number of claims at acceptance: 36

Applicant and inventor details

Applicant name(s) and address(es) *(as it will appear on certificate/s)* :

4D Molecular Therapeutics Inc. of 5980 Horton Street, Suite 460 Emeryville CA 94608 United States of America

Inventor name(s): Kirn, David H.
Kotterman, Melissa
Schaffer, David

Agent details

Agent Name WRAYS PTY LTD
Address for correspondence: L7
863 Hay St
Perth WA 6000
Australia
Address for legal service: L7
863 Hay St
Perth WA 6000
Australia

Prior art details

Prior art documents:

WO 2018/022905 A2
WO 2017/218842 A1

Priority details

Number	Date	Country
62/384,590	7 September 2016	US
62/454,612	3 February 2017	US
62/336,441	13 May 2016	US
62/378,106	22 August 2016	US

Your reference 274011
Application number 2017261812

International classification

C07K 14/075 (2006.01)
A61K 48/00 (2006.01)
C12N 15/62 (2006.01)
C12N 15/63 (2006.01)
C12N 15/861 (2006.01)
C12Q 1/70 (2006.01)

Relevant dates

Acceptance date: 5 April 2019
Acceptance to be advertised: 18 April 2019
Complete filing date: 12 May 2017
OPI date: 16 November 2017
Continuation fee due date: 12 May 2021
Date of patent: 12 May 2017
Expiry date: 12 May 2037
National phase entry date: 26 October 2018

[En papel membrete del gobierno de Australia – Propiedad Industrial de Australia]

9 de abril de 2019

NOTIFICACIÓN DE ACEPTACIÓN DE SOLICITUD DE PATENTE

WRAYS PTY LTD

L7, 863 Hay St, Perth WA 6000, Australia

Su referencia: 274011

No. de solicitud: 2017261812

Solicitante: 4D Molecular Therapeutics Inc.

Estimados WRAYS PTY LTD:

La presente solicitud de patente fue aceptada el 5 de abril de 2019. La memoria descriptiva aceptada incorpora las siguientes modificaciones:

Modificación según el Artículo 104 incluido el número de ítem: 3

Cantidad de reivindicaciones al momento de la aceptación: 36

La notificación de la aceptación se publicará en el Boletín Oficial de patentes de Australia del 18 de abril de 2019.

Les solicitamos tengan a bien revisar los datos adjuntos para asegurarnos de que son correctos. Si desean efectuar alguna corrección antes de la concesión, deberán realizarla dentro de los tres meses posteriores a la publicación de la aceptación.

Se les emitirá una invitación de pago del arancel de aceptación. Este arancel incluirá un ítem adicional si la cantidad de reivindicaciones excede las 20 reivindicaciones.

Los datos relativos a esta solicitud de patente se encuentran disponibles en AusPat, nuestra base de datos de patentes de Australia.

Atentamente.

IP Australia

Los datos consignados en la presente relacionados con la solicitud número 2017261812 están actualizados al 10 de abril de 2019 23:29 AEST.

RESUMEN DE LA ACEPTACIÓN

Datos generales de patente estándar:

Número de patente: 2017261812

Título: CÁPSIDES VARIANTES DE VIRUS ADENOASOCIADOS Y MÉTODOS DE USO DE ÉSTAS.

Su referencia: 274011

Cantidad de reivindicaciones al momento de la aceptación: 36

Datos del solicitante y de los inventores:

Nombre y domicilio del solicitante *(como va a figurar en el título de la patente)*

4D Molecular Therapeutics Inc. of Horton Street, Suite 460, Emeryville CA 94608, Estados Unidos de Norteamérica

Nombres de los inventores: David H. Kirn; Melissa Kotterman; David Schaffer

Datos del agente

Nombre del agente: WRAYS PTY LTD

Domicilio postal: L7, 863 Hay St., Perth WA 6000, Australia

Domicilio para notificaciones: L7, 863 Hay St., Perth WA 6000, Australia

Arte Previo:

Documentos del arte previo:

WO 2018/022905 A2

WO 2017/218842 A1

Prioridades

Número	Fecha	País
62/384,590	7 de septiembre de 2016	EE. UU.
62/454,612	3 de febrero de 2017	EE. UU.
62/336,441	13 de mayo de 2016	EE. UU.
62/378,106	22 de agosto de 2016	EE. UU.

Clasificación internacional

C07K 14/075 (2006.01)

A61K 48/00 (2006.01)

C12N 15/62 (2006.01)

C12N 15/63 (2006.01)

C12N 15/861 (2006.01)

C12Q 1/70 (2006.01)

Fechas importantes

Fecha de aceptación: 5 de abril de 2019

Publicación de la aceptación: 18 de abril de 2019

Fecha de presentación: 12 de mayo de 2017

Fecha de apertura al público: 16 de noviembre de 2017

Vencimiento de la primera anualidad: 12 de mayo de 2021

Fecha de la patente: 12 de mayo de 2017

Fecha de vencimiento: 12 de mayo de 2037

Fecha de entrada en fase nacional: 26 de octubre de 2018

4

An Evolved Adeno-associated Viral Variant Enhances Gene Delivery and Gene Targeting in Neural Stem Cells

Jae-Hyung Jang¹⁻⁴, James T Koerber¹⁻³, Jung-Suk Kim⁴, Prashanth Asuri¹⁻³, Tandis Vazin¹⁻³, Melissa Bartel¹⁻³, Albert Keung¹⁻³, Inchan Kwon¹⁻³, Kook In Park^{5,6} and David V Schaffer¹⁻³

¹Department of Chemical Biomolecular Engineering, University of California, Berkeley, California, USA; ²Department of Bioengineering, University of California, Berkeley, California, USA; ³The Helen Wills Neuroscience Institute, University of California, Berkeley, California, USA; ⁴Department of Chemical and Biomolecular Engineering, Yonsei University, Seoul, Korea; ⁵Department of Pediatrics, Yonsei University College of Medicine, Seoul, Korea; ⁶BK21 Project for Medical Science, Yonsei University College of Medicine, Seoul, Korea

Gene delivery to, and gene targeting in, stem cells would be a highly enabling technology for basic science and biomedical application. Adeno-associated viral (AAV) vectors have demonstrated the capacity for efficient delivery to numerous cells, but their application to stem cells has been limited by low transduction efficiency. Due to their considerable advantages, however, engineering AAV delivery systems to enhance gene delivery to stem cells may have an impact in stem cell biology and therapy. Therefore, using several diverse AAV capsid libraries—including randomly mutagenized, DNA shuffled, and random peptide insertion variants—we applied directed evolution to create a “designer” AAV vector with enhanced delivery efficiency for neural stem cells (NSCs). A novel AAV variant, carrying an insertion of a selected peptide sequence on the surface of the threefold spike within the heparin-binding site, emerged from this evolution. Importantly, this evolved AAV variant mediated efficient gene delivery to rat, mouse, and human NSCs, as well as efficient gene targeting within adult NSCs, and it is thus promising for applications ranging from basic stem cell biology to clinical translation.

Received 14 August 2010; accepted 26 November 2010; published online 11 January 2011. doi:10.1038/mt.2010.287

INTRODUCTION

Stem cells—defined by the hallmark properties of self-renewal, i.e., the capacity to expand, maintain themselves in an undifferentiated state, as well as differentiation into one or more specialized lineages—have considerable biomedical potential. However, better approaches must be developed to control their behavior. Gene delivery can serve as a valuable tool for numerous applications, including modulating stem cell fates by up-/down-regulating specific genes, developing stem cell-based models of human disease, and tracking of stem cells or their progeny upon implantation *in vivo*.¹⁻²

Current methods to mediate gene delivery to stem cells include plasmid- and viral vector-mediated delivery. Plasmid-based gene transfer offers typically transient gene expression and a large carrying capacity;³ however, current plasmid delivery systems for stem cells can suffer from low transfection efficiency.⁴ Alternatively, viral vectors based on retroviruses and lentiviruses offer high infectivity and stable gene expression and have been extensively employed in stem cell studies for genetic marking, protein or RNA overexpression, and targeted knock-down of endogenous gene expression.^{5,6} However, in some cases transient expression is desirable, such as for the temporary overexpression of regulatory signals to manipulate stem cells.⁷⁻⁹ Furthermore, insertional mutagenesis is a concern for potential downstream clinical application of retroviruses.¹⁰ Adenoviral vectors also have potential for efficient delivery and transient gene expression in stem cells,¹¹ though the most commonly used adenoviral vectors carry numerous virally-encoded factors that may perturb cell function.¹² An efficient vector that mediates transient gene expression in stem cells may therefore have utility in the field.

In addition to gene delivery, the capacity for gene targeting via homologous recombination at specific loci in a mammalian stem cell genome has many potential applications, including basic investigation of processes of organismal development and mechanisms of human disease, the targeted introduction of genetic material into a genome without insertional mutagenesis, and therapeutic gene correction.¹³ However, plasmid-mediated gene targeting is typically very inefficient. Recent approaches that use zinc finger nucleases to enhance gene targeting efficiencies are highly promising, though they require engineering a new zinc finger nuclease for each new locus one would like to manipulate.^{14,15}

Adeno-associated virus (AAV) is a parvovirus with a 4.7 kb single-stranded DNA genome containing two genes (*rep* and *cap*), and vectors based on AAV have proven safe and effective for clinical application.¹⁶⁻¹⁸ Exposure to AAV results in persistent, latent infection in a broad range of both dividing and nondividing cells.^{19,20} Additionally, Russell and colleagues have demonstrated that AAV vector genomes can mediate homologous recombination

Correspondence: David V Schaffer, Department of Chemical and Biomolecular Engineering, University of California, Berkeley, 274 Stanley Hall, Berkeley, California 94720-1462, USA. E-mail: schaffer@berkeley.edu

with target sequences in a cellular genome at efficiencies 10^3 – 10^4 -fold higher than plasmid constructs.²¹ This capability has been successfully applied to cells that AAV can effectively transduce, including fibroblasts and recently mesenchymal stem cells, which has led to the development of new animal models of human disease.^{22,23} Although AAV-mediated gene expression and targeting would be effective tools for stem cell studies, AAV vectors are unfortunately extremely inefficient for gene delivery numerous stem cell types.^{24,25} Therefore, the development of new AAV variants capable of efficient stem cell infection may have impact in stem cell biology and regenerative medicine.

Embryonic neural stem cells (NSCs) are centrally involved in the development of the nervous system,²⁶ and adult NSCs play a fundamental role in learning and memory and may yield new therapies for neurodegenerative disorders.²⁷ However, AAV is extremely inefficient on NSCs.²⁵ Our recent work highlights the potential of directed evolution to create new viruses with enhanced gene delivery properties.^{18,28–31} In this approach, we generate large ($\sim 10^7$) and highly diverse libraries of viral variants, and the application of an evolutionary pressure—such as the ability to infect a stem cell—selects for the “fittest” variants. Importantly, this general approach can yield successful variants in the absence of mechanistic knowledge of which step(s) in the viral entry pathway—such as cell surface binding, intracellular trafficking, and/or nuclear entry—pose(s) the most substantial barrier to infection. In this study, we have applied directed evolution to create a novel AAV vector with the capacity for high efficiency gene delivery to adult rat NSCs, and this variant was also efficient in murine and human NSCs. Furthermore, this high delivery efficiency was applied to enhance gene correction frequencies in NSCs, potentially providing a tool with potentially strong utility for both basic biology and biomedical application.

RESULTS

AAV library generation

We found that several wild-type AAV serotypes (2, 4, 5, 6, and 8) exhibited low transduction efficiencies [0.7–9.7% with multiplicity of infection (MOI) 10^5] of NSCs (Figure 1a), as previously described.^{24,25} These results motivated pursuing directed evolution for the development of novel AAV variants that can more effectively infect NSCs. Three large ($>10^6$ independent clones each), diverse libraries were used: (i) random viral chimeras created via pairwise DNA shuffling of the AAV2 *cap* gene with that of AAV4, 5, 6, or 8, (ii) AAV2 capsid variants with a random 7 amino acid sequence inserted into a key loop region,³² and (iii) random point mutants of the *cap* gene of AAV2 generated via error prone PCR¹⁸ (Figure 1b). Libraries were packaged as previously described to generate a large pool of AAV mutant capsids containing the genome encoding those mutants.¹⁸ NSCs, isolated from the hippocampus of adult Fisher 344 rats,²⁷ were infected by each AAV library, adenovirus serotype 5 was added to induce the replication and rescue of the “successful” AAV variants, the surviving AAV *cap* genes were recovered by PCR and cloned into the AAV genome plasmid, and the next generation library was packaged. After four such iterative selection steps, the resulting viral pool was subjected to additional mutagenesis by error prone PCR on the selected *cap* genes, followed by further iterative selections

(Figure 1b). During the three rounds of evolution and 10 selection steps, the variant pool became progressively more infectious than wild-type AAV2 (data not shown).

AAV variants selected through directed evolution

DNA sequencing analysis revealed that the final pools contained seven AAV variants: two shuffled, three peptide-inserted, and two error prone clones (Figure 2, Table S1). Interestingly, the shuffled clones (r3.1, r3.10) were composed of AAV2 and AAV5 sequences, with additional point mutations located on the viral surface. Although AAV2 and 5 were less efficient than AAV4 and AAV6 (Figure 1a), the low efficiencies of AAV4 and AAV6 [$<10\%$ green fluorescent protein positive (GFP⁺) NSCs at a MOI 10^5] may be insufficient to enable wild-type *cap4* and *cap6* to be represented in the selected pool. Furthermore, combinations of fragments of *cap4* and *cap6* with *cap2* likely yield gene delivery properties that are distinct from the parental *cap4* and *cap6*. In addition, two peptide insertions were present in three clones (r3.3, r3.21, r3.45), which also carried several point mutations.

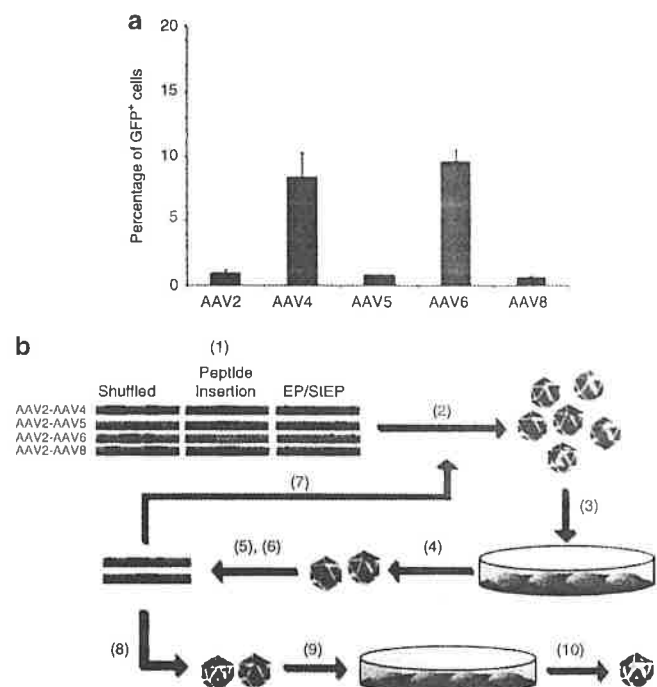


Figure 1 Rationale and strategy for engineering novel AAV variants. **(a)** Rat neural stem cell (NSC) infection with wild-type adeno-associated viral (AAV) 2, 4, 5, 6, and 8. Cells were infected by addition of green fluorescent protein (GFP) vectors at a multiplicity of infection (MOI) of 10^5 . Cells were incubated with virus for 24 hours, followed by analysis of GFP expression by flow cytometry after an additional 24 hours. Even at this high MOI, the highest gene delivery efficiency was $<10\%$. **(b)** Directed evolution of novel AAV vectors for enhanced NSC infection: (1) plasmid libraries were generated by DNA shuffling, peptide insertion, and error prone PCR [error prone (EP)], (2) viral libraries were packaged, (3) NSCs were infected with libraries, followed by adenovirus addition, (4) the resulting amplified AAV variants were harvested, and (5) viral genomes were isolated and cloned. The cycle of steps (2)–(5) was iterated (7) with (6) additional EP PCR after the fourth and seventh cycles. Newly evolved clones were used to produce recombinant virus (8) and analyzed for the ability to infect NSCs (9), and the final clone was selected (10).

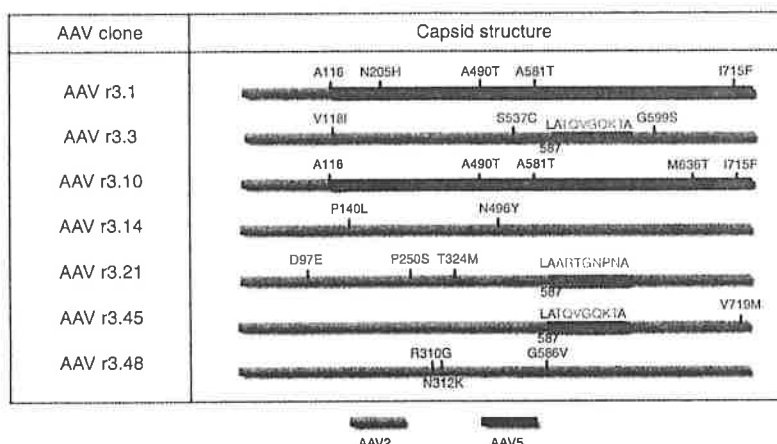


Figure 2 Schematic maps of novel adeno-associated viral (AAV) variants. AAV r3.1 denotes the first AAV clone isolated after three rounds of evolution. The gray and blue coloring depicts AAV2 capsid and AAV5 capsid regions, respectively. In the peptide-inserted vectors, "LA" and "A" were inserted as linkers.

Novel AAV variant with enhanced properties for NSC infection

These capsid mutants were used to package recombinant vector encoding GFP driven by a cytomegalovirus (CMV) promoter, which revealed that six of the variants exhibited similar or slightly improved efficiencies compared to AAV2 or AAV5 (Supplementary Figure S1). Additional rounds of selection would likely have eliminated these variants to yield a dominant clone.³⁰ In stark contrast, however, the variant AAV r3.45 mediated high delivery efficiencies (Figure 3a–f). At MOI of 10^3 , 10^4 , and 10^5 , AAV r3.45 infected 14.4–50 fold more cells relative to AAV2 and AAV5 (Figure 3d–f). Importantly, 7 days after infection with AAV r3.45, nearly 100% of GFP⁺ cells maintained expression of the neural progenitor marker nestin (Figure 4a), indicating that infected cells remained in an immature state.

AAV r3.45 carried an insertion of LATQVGQKTA (where LA and A are linker residues) at amino acid 587 with an additional V719M mutation (Figure 4b). The inserted peptide likely lies on the surface of the threefold spike within the heparin-binding site (Figure 4c,d), and the point mutation is also surface-exposed. AAV r3.3, which harbored the same peptide insertion but with three different point mutations (*i.e.*, V118I, S537C, and G599S), exhibited considerably lower packaging titers and infectivities than AAVr3.45, indicating that slight capsid modifications can significantly impact viral properties (Figure 1 and Supplementary Figure S1). Importantly, AAV r3.45 was additionally highly infectious for both murine and human NSCs (Figure 5a–c). It was also infectious in several other nonpermissive cell types (Supplementary Figure S2), though not in human or murine embryonic stem cells (Supplementary Figure S3a,b), indicating that its benefits are somewhat cell selective. However, despite its lack of cell specificity, its enhanced gene delivery to cultured NSCs from three different species met the goals of this study and indicates that AAV r3.45 may be used in a variety of stem cell applications, such as transient expression of factors or gene targeting to generate knock-ins/-outs for basic investigation or the development of disease models.

To investigate the roles of the peptide insertion and the V719M mutation for the enhancement of NSC infection, the V719M mutation was reversed via site-directed mutagenesis, and the resulting variant, which contains the peptide insertion without the V719 mutation, was packaged and employed to infect NSCs. This correction of the V719M mutation did not influence the infectivity, indicating the peptide sequence displayed on the AAV r3.45 capsid is critical to enhance gene delivery to NSCs (Supplementary Figure S4). However, although a BLAST search (<http://www.ncbi.nlm.nih.gov/BLAST/>) revealed that the inserted sequence bears sequence similarity to a number of intracellular enzymes (*e.g.*, *Halobacterium salinarum* naphthoate synthase), it was not similar to any secreted mammalian or viral capsid proteins.

To determine whether this variant may still use heparan sulfate as its primary receptor, despite its peptide insertion within the AAV2 heparin-binding domain,³³ we compared the heparin affinity of AAV r3.45 to that of wild-type AAV2. Purified AAV2 or AAV r3.45 was loaded onto a heparin column and eluted with a series of increasing NaCl concentration (150–750 mmol/l, and 1 mol/l), and a portion of each fraction was added to 293T cells to quantify infectious virus (Supplementary Figure S5). Similar to previous reports,^{18,29} fractions from 450 to 550 mmol/l NaCl contained the majority of wild-type AAV2. However, AAV r3.45 eluted in a sharp peak between 350 mmol/l and 450 mmol/l, indicating that heparan sulfate may still be primary receptor, but its affinity to heparan sulfate may be slightly altered.

Enhanced gene correction frequency with newly designed AAV variant

We next assessed the capacity of AAV r3.45 to mediate gene correction in NSCs. GFP cDNA harboring a stop codon (mutGFP1) was introduced into adult rat NSCs using a lentiviral vector. In parallel, a promoter-less GFP with a distinct stop codon (mutGFP2) was packaged into a recombinant AAV vector (Figure 6), and the resulting gene targeting construct carrying mutGFP2 had 586 nucleotides of homology upstream and

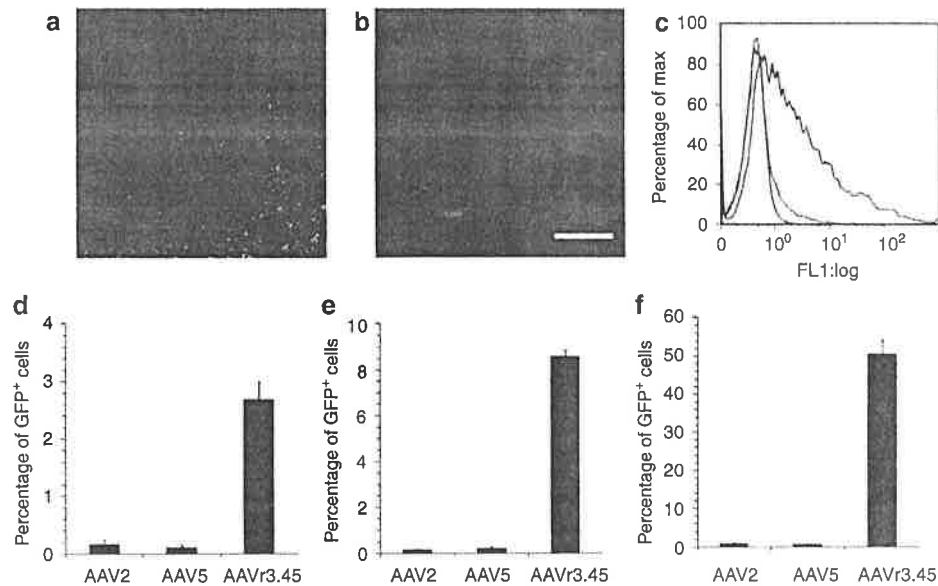


Figure 3 Neural stem cell (NSC) transduction by adeno-associated virus (AAV) r3.45. (a) Phase-contrast microscopy and (b) fluorescence microscopy of cells infected by AAV r3.45 [genomic multiplicity of infection (MOI): 10⁵] shows a high level of transduction. Bar = 100 μ m. (c) Histogram of green fluorescent protein (GFP)-expressing cells: MOI 10³ (red), MOI 10⁴ (green), MOI 10⁵ (blue). Efficiencies of recombinant AAV2, AAV5, and AAV r3.45 were quantified by flow cytometry at (d) MOI 10³, (e) 10⁴, and (f) 10⁵.

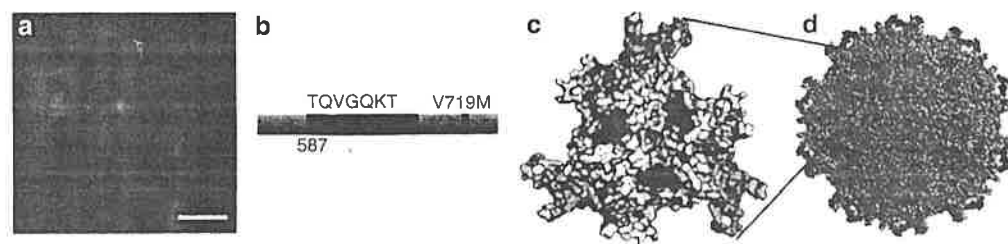


Figure 4 Neural stem cells (NSCs) infected with adeno-associated virus (AAV) r3.45 remain undifferentiated. (a) Seven days after infection with AAV r3.45, immunostaining revealed that the majority of green fluorescent protein expressing cells were positive for nestin, a neural progenitor cell marker. Bar = 20 μ m. (b) Schematic representation of the peptide sequence, point mutation, and heparin-binding residues of AAV clone r3.45. (c) Structural model of a VP3 subunit of AAV r3.45, as well as (d) the full capsid. The color scheme for panels b–d is: V719M (blue), peptide sequence at amino acid 587 (red), and heparin-binding residues (R585/R598) (purple).

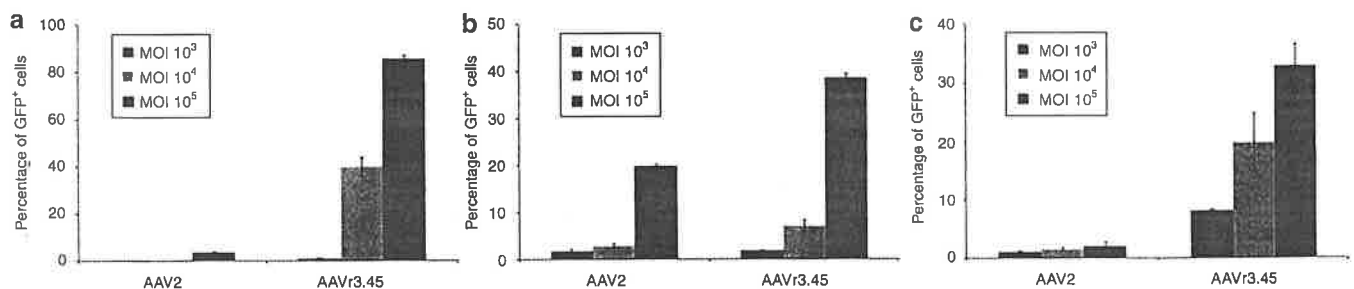


Figure 5 Enhanced neural stem cell tropism of AAV r3.45. Comparison of transduction efficiencies of adeno-associated virus (AAV) r3.45 versus AAV2 for (a) murine neural stem cells (NSCs), (b) human fetal NSCs, and (c) human neural progenitor cells. Cells were infected at a multiplicity of infection (MOI) of 10³, 10⁴, and 10⁵, and green fluorescent protein (GFP) expression was analyzed via flow cytometry 48 hours after infection.

1,548 nucleotides downstream of the amber mutation in mutGFP1. Upon infection of NSC-mutGFP1 cells with the targeting vector, recombination with the integrated lentiviral vector corrected the amber mutation in mutGFP1, resulting in fluorescent NSCs (Figure 7a). Importantly, AAV r3.45 demonstrated a significantly improved capacity (six- to tenfold) to repair the

single-base pair mutation in NSC-mutGFP1 cell line compared to AAV2 and AAV5 (Figure 7b, Table S2). Controls, including the infection of NSC-mutGFP1 cell lines by AAV CMV Luc (Figure 7c) and naive NSCs infected by the targeting vectors (data not shown), did not yield GFP expression, verifying the specificity of the process.

Verification of gene correction

To confirm the GFP gene correction, GFP⁺ cells were isolated by fluorescence-activated cell sorting 14 days after AAV infection, and cellular genomic DNA was subjected to PCR to amplify the GFP sequence specifically from the lentiviral vector. Sequence analysis of 24 cloned PCR fragments revealed 19 corrected

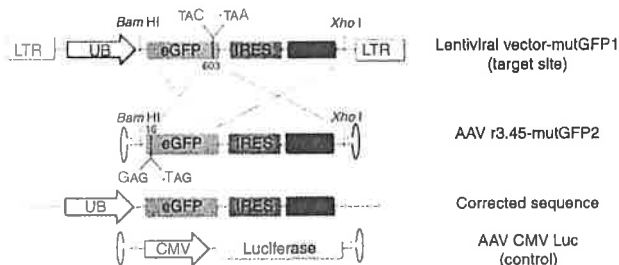


Figure 6 Schematic representation of a lentiviral vector carrying UB-mutGFP1-IRES-PuroR, the adeno-associated viral (AAV)-mutGFP2 targeting vector, AAV cytomegalovirus (CMV) promoter Luc, and the corrected sequence at the target site. Genetic elements include: the ubiquitin promoter (UB), internal ribosome entry site (IRES), puromycin acetyltransferase (PuroR), mutGFP1 [a single-base substitution (C → A) 603 nucleotides (nt) from the 5' end of green fluorescent protein (GFP)], and mutGFP2 [a single point mutation (G → T) 16 nt into GFP]. eGFP, enhanced green fluorescent protein; LTR, long terminal repeat.

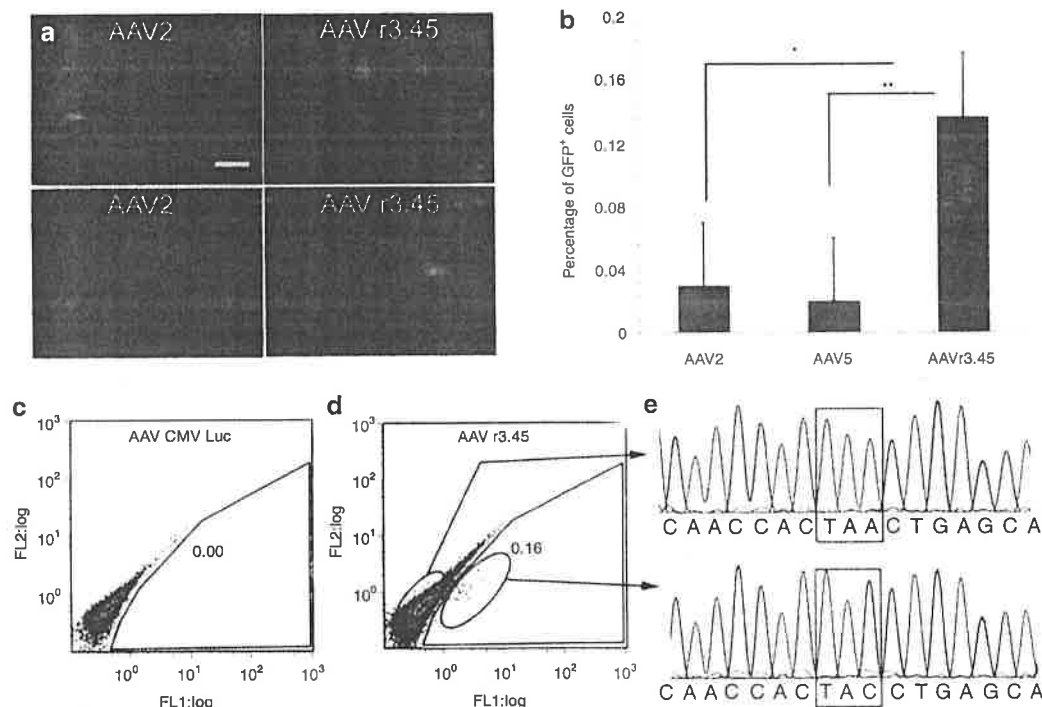


Figure 7 Adeno-associated virus (AAV) r3.45-mediated gene correction by homologous recombination with a nonfunctional green fluorescent protein (GFP) expressed in neural stem cells (NSC) and verification of gene correction. **(a)** Immunostaining of GFP⁺, gene-corrected NSC-mutGFP1 cells 72-hours postinfection with the AAV-mutGFP2 targeting vectors: GFP (green), DAPI (blue). Example images with a lower (top row) and higher (bottom row) proportion of cells with a corrected GFP are shown. Bar = 100 μm. **(b)** Gene targeting frequencies quantified by flow cytometry 3-days postinfection. *, **Statistical differences of $P < 0.05$ and $P < 0.001$, respectively. **(c)** Flow cytometry analysis of NSC-mutGFP1 cells infected by AAV cytomegalovirus (CMV) promoter Luc as a negative control. **(d)** Fluorescence-activated cell sorting analysis of gene targeting in NSC-mutGFP1 cells 3-days postinfection with AAV r3.45-mutGFP2, which resulted in GFP⁺ cells shown within the red circled regions. **(e)** Sequence chromatographs of "uncorrected" (top panel) and "corrected" (bottom panel) GFP DNA fragments amplified from sorted, GFP⁺ NSC-mutGFP1. Genomic DNA from FL1-positive population was isolated, subjected to PCR, and cloned using StrataClone PCR cloning kit (Stratagene). Nineteen of 24 clones from cells in the GFP⁺ population showed correction from an A to a C.

sequences (79.2%), with the five uncorrected sequences presumably resulting from NSCs infected with multiple lentiviral vectors (**Figure 7d,e**). In parallel, sorted cells were expanded for an additional 26 days (total 40 days after AAV infection), and ~94.5% of cells maintained GFP and nestin expression (**Figure 8**).

DISCUSSION

AAV is a highly versatile vector with a broad-host range; however, its delivery efficiency to certain cell types, including neural or embryonic stem cells, is low.^{24,25} One potential reason is a lack of necessary cell surface receptors for a given serotype (e.g., heparan sulfate, integrins), resulting in low attachment affinity and cellular internalization. Alternatively, intracellular trafficking pathways can also pose barriers, as for example nuclear trafficking of AAV is a key rate-limiting step for AAV transduction of some cell types such as murine fibroblasts.³⁴ Additionally, noncapsid related limitations including double-stranded DNA synthesis can compromise AAV-mediated gene expression.³⁵ Our finding that a short peptide insertion into the exterior surface of AAV2 can substantially increase delivery efficiency implicates capsid-related steps such as receptor binding or intracellular trafficking as potential barriers.

The slight shift in the heparin affinity of this variant AAV r3.45 (**Supplementary Figure S5**) may correspond to a significant

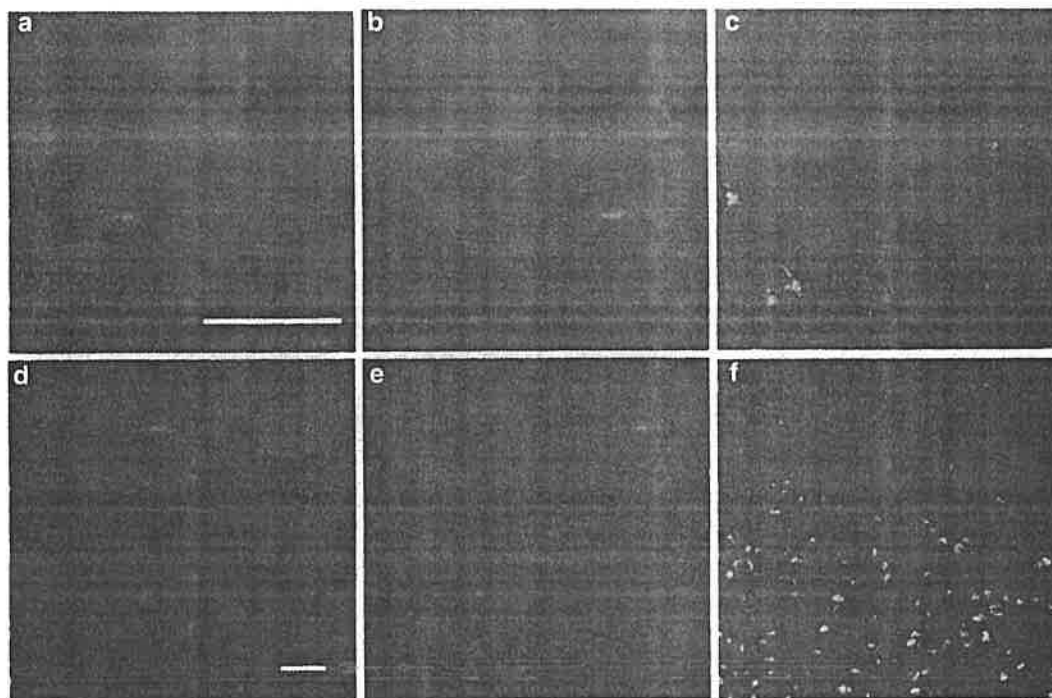


Figure 8 Maintenance of the undifferentiated state of gene-corrected, green fluorescent protein (GFP)-expressing neural stem cells 35-days postinfection. (Nestin: red, GFP: green). The nuclei of cells were co-stained with TOPRO3 (blue). Bar = 50 μ m. (a–c: $\times 200$, d–f: $\times 50$, a,d: GFP, b,e: nestin/TOPRO3, c,f: GFP/nestin/TOPRO3).

modulation of heparan sulfate binding, potentially due to the inserted peptides' conformational alteration of a key exposed loop of the heparin binding domain. However, the fact that the majority of AAV r3.45 still eluted at relatively high NaCl concentrations indicates that this variant may still utilize heparan sulfate as its primary receptor. In addition, the LATQVGQKTA peptide may thus contribute to binding to an alternate secondary receptor or to an intracellular transduction step. Future studies may elucidate this mechanism of action.

The central goal of the directed evolution in this study was to create novel vectors that can better infect NSCs; therefore, positive selection was applied for enhanced adult rat NSC transduction, and negative selection was not applied to remove tropism for other cell types. The approach successfully yielded a novel AAV variant with significantly enhanced infectivity of not only adult rat but also mouse and two human NSCs. As anticipated this enhanced infectivity was not accompanied by specificity for NSCs, as AAV r3.45 was more efficient than AAV2 for several cell types, including primary human astrocytes (Supplementary Figure S2). Accordingly, *in vivo* AAV r3.45 would be anticipated to transduce NSCs as well as other surrounding cell types (*i.e.*, astrocytes, oligodendrocytes). However, alternative selection strategies could be implemented to achieve specificity of delivery.

The development of an efficient vehicle for transient gene expression in neural stem cells has applications ranging from the overexpression of factors to induce the differentiation or "programming" of neural stem cells, as well as potentially expression of defined transcription factors to reprogram cells.^{36,37} Moreover, enhanced transduction and gene targeting capabilities in NSCs may facilitate stem cell biology and regenerative medicine

investigations, potentially including the introduction of defined mutations for basic studies, generation of reporter cell lines, and therapeutic gene correction. By yielding a variant with high efficiency gene delivery and gene targeting in NSCs, directed evolution is further established as a powerful approach to engineer novel capabilities into AAV vectors.^{18,28–31} Further reverse engineering analysis of the gene delivery mechanisms of this variant may elucidate mechanisms that limit gene delivery in stem cells, as well as the other cell types efficiently infected by this variant. Finally, this general approach may in the future be applied to other cell and stem cell types, thereby extending the utility of AAV to additional biotechnology and biomedical applications.

MATERIALS AND METHODS

NSC culture. Adult rat NSCs, isolated from the hippocampal region of 6-week-old female Fisher 344 rats as described,^{27,38} were cultured on tissue culture polystyrene dishes previously coated with poly-ornithine (10 μ g/ml) (Sigma-Aldrich, St Louis, MO) and mouse laminin (5 μ g/ml) (Invitrogen, Carlsbad, CA), in Dulbecco's modified Eagle's medium (DMEM)/F-12 (1:1) (Invitrogen) containing N-2 supplement (Invitrogen) and 20 ng/ml of recombinant human fibroblast growth factor 2 (Peprotech, Rocky Hill, NJ). Under these conditions, nearly 100% of cells express nestin and Sox2 (data not shown). Every 2–3 days (>60% confluency), cells were split into a fresh dish by mechanical dissociation using polished pasteur pipettes. Adult murine NSCs (a kind gift of Prof Andrew Wurmser, University of California at Berkeley) were cultured in DMEM with 20 ng/ml of recombinant fibroblast growth factor 2 (Peprotech) and 8 μ g/ml of heparin. Human fetal NSCs (13 weeks of gestational age) were derived from the telencephalon (HFT13) as previously described³⁹ and were cultured in DMEM/F-12 containing N-2 supplement, 20 ng/ml of fibroblast growth factor 2, 8 μ g/ml of heparin (Sigma-Aldrich), and 10 ng/ml of leukemia-inhibitory factor (Chemicon, Temecula, CA). Human neural progenitor

cells were derived from a variant of the human embryonic stem cell line BG01 (BresaGen, Athens, GA), BG01V2, using the stromal derived inducing activity involving a co-culture system of human embryonic stem cells and mouse stromal cell line PA6 (Riken BioResource Center Cell Bank, Tsukuba, Japan) (Vazin *et al.* manuscript in preparation) and were cultured in DMEM/F-12 supplemented with N-2, 20 ng/ml basic fibroblast growth factor, 20 ng/ml epidermal growth factor (Invitrogen), and 50 U/ml Penicillin/Streptomycin.

Plasmid library construction. Random mutagenesis libraries were generated by subjecting *cap* genes from AAV2 to error prone PCR as previously described.¹⁸ Peptide display libraries were generated similar to previous reports.^{32,40} Briefly, a unique *Avr* II was introduced into pSub2Cap2 between amino acid 587 and 588 by PCR mutagenesis. A random 21 base pair oligonucleotide [5'-GGAGGGCTAGCA (NNK), GCTAGCAAAAGCGGGGGAGAGTGAGG-3'] was annealed to an antisense primer 5'-CCTCACTCTCCCCCGCT-3', followed by synthesis of the second strand. The resulting dsDNA inserts were digested with *Nhe* I and ligated into the unique *Avr* II site of pSub2Cap2. Pairwise DNA shuffling libraries of AAV2 with AAV4, AAV5, AAV6, or AAV8 (AAV2-AAV4, AAV2-AAV5, AAV2-AAV6, and AAV2-AAV8) were generated by PCR amplification of *cap* genes from AAV2, AAV4, AAV5, AAV6, and AAV8 and followed by insertion into pSub2. Each viral gene was then amplified by PCR, and DNA shuffling was performed as previously described.^{41,42} Briefly, DNase I digestion of equimolar amounts of PCR-amplified *cap* genes yielded fragments ranging from 50 to 500 base pair in size. Fragments were gel purified and reassembled without primers in the following conditions: 96°C, 3 minutes; 40 cycles of 94°C, 55°C, and 72°C; and 72°C, 10 minutes. Assembled fragments were further amplified by PCR and cloned into pSub2 for replication competent AAV production.

Selection of viral libraries. The plasmid library was used to package replication competent AAV as previously described,^{18,43} and the resulting viral libraries were then harvested as described.⁴⁴ Virus was purified using OptiPrep (60%, wt/vol iodixanol) (Sigma-Aldrich) density gradient ultracentrifugation according to the manufacturer's instructions, and genomic titers of AAV viral libraries were determined by quantitative real time PCR on a Bio-Rad iCycler (Bio-Rad, Hercules, CA) using a Taqman probe (Biosearch Technologies, Novato, CA).¹⁸

NSCs, seeded at a density of 6×10^5 per well in 6-well tissue culture plates coated with poly-ornithine/laminin one day prior to infection, were infected with either wild-type AAV2 or the AAV libraries at genomic MOIs of 10, 100, and 1,000. After a 24-hour exposure, cells were co-infected with adenovirus serotype 5 in fresh medium (*i.e.*, DMEM/F12 + N2, fibroblast growth factor 2). Once cells exhibited cytopathic effect, after approximately two additional days of culture, they were lysed by three freeze/thaw steps, followed by the treatment with benzonase (1 unit/ml) (Sigma-Aldrich) at 37°C for 30 minutes. Adenovirus was inactivated by incubating the viral lysates at 50°C for 30 minutes. To quantify viral libraries rescued from the each selection, viral genomic DNA was extracted from DNase-resistant particles. At each step, the total viral genomic DNA rescued was compared for the viral libraries versus wild-type AAV2.

Evolution of viral libraries. One round of evolution is defined as *cap* mutagenesis followed by several selection steps, *i.e.*, NSC infection and adenovirus rescue. After 3–4 selection steps, the rescued AAV *cap* genes were randomly mutated by error prone PCR as previously described.¹⁸ The resulting PCR products were digested with *Hind* III and *Not* I, and the fragments were ligated into pSub2 to create a new replication competent AAV plasmid library with at least 10^6 independent clones. Viral library packaging was then conducted by the calcium phosphate transient transfection method,¹⁸ and viral titration and purification were performed as with the initial library construction.

After the evolution was completed, AAV *cap* genes were extracted from the final pool of successful AAV variants, and genomic sequences

were analyzed at the UC Berkeley DNA Sequencing Facility. The extracted viral genomes were amplified by PCR and inserted into the pXX2 *Not* I recombinant AAV (rAAV) packaging plasmid.^{22,33} Three-dimensional models of the VP3 subunit were generated using Swiss Model (<http://swissmodel.expasy.org>) with the coordinates of AAV2 (Protein Databank accession no. 1LP3) supplied as template, and images were rendered in Pymol (<http://pymol.sourceforge.net/>) and Rasmol.

Heparin column chromatography. AAV r3.45 and AAV2 binding to heparin were analyzed as previously described.^{18,29} Briefly, $\sim 10^{11}$ purified genomic particles of virus were loaded onto a 1 ml HiTrap heparin column (GE Healthcare Life Sciences, Pittsburgh, PA), which was previously equilibrated with Tris buffer (50 mmol/l, pH 7.5) containing 150 mmol/l NaCl. Elutions were performed with 0.75 ml of Tris buffer containing increasing increments of 50 mmol/l NaCl up to 750 mmol/l, followed by a 1 ml/l NaCl wash. Small fractions (75 μ l) of each elution were added to 2.5×10^5 293T cells cultured in 12-well dishes and employed to quantify the level of infectious virus present. Forty-eight hour postinfection, GFP expression was measured with Becton Dickinson fluorescence-activated cell sorting Caliber (Yonsei University College of Medicine Medical Research Center, Seoul, Korea).

NSC transduction. To test the infectivity of rAAV *cap* clones, vectors carrying cDNA encoding GFP were packaged using the transient transfection method. Briefly, an equal mass (17 μ g) of three plasmids—a AAV helper plasmid (carrying a *cap* variant), CMV GFP vector plasmid containing ITR (pAAV CMV GFP SN), and pHelper plasmid—were transfected via calcium phosphate into AAV293 cells,^{18,27} and viral vectors were harvested and purified as described above. Genomic titers of each rAAV vector were determined by quantitative PCR.

One day prior to AAV infection, NSCs were plated onto 24-well dishes at a density of 30,000 cells/well. On day 1, cells were infected by the rAAV clones at genomic MOIs (10^3 , 10^4 , and 10^5) and incubated for additional 48 hours before analysis. The resulting transduction efficiencies of the individual rAAV variants on rat and murine NSCs were quantified with a Beckman-Coulter EPICS XL flow cytometer (UC Berkeley Cancer Center, Berkeley, CA) and compared to those of AAV2 and AAV5, which were chosen as controls since the two shuffled clones that emerged from the evolution carried fragments from AAV *cap2* and AAV *cap5* (Figure 1). Human NSC and neural progenitor cell transduction efficiencies were quantified with a Becton Dickinson fluorescence-activated cell sorting Caliber (Yonsei University College of Medicine Medical Research Center) and Beckman-Coulter FC 500 flow cytometer (Berkeley Stem Cell Center, Berkeley, CA), respectively.

Immunofluorescence staining. To analyze AAV r3.45 infection of undifferentiated NSCs, and the effects of AAV infection on NSC differentiation, immunostaining for nestin (undifferentiated NSCs), β -III-tubulin (neurons), and glial fibrillary acidic protein (astrocytes) was conducted.²¹ NSCs were seeded onto 8-well chamber slides and infected as indicated above. At 48 hours postinfection, cells were fixed in 4% paraformaldehyde for 15 minutes and blocked with 5% goat serum in 0.3% Triton X-100/phosphate-buffered solution for 30 minutes at room temperature. Cells were incubated overnight at 4°C with primary antibodies. For nestin/ β -III-tubulin co-staining, rabbit anti-rat nestin (1:500 dilution; Abcam) and mouse anti- β -III-tubulin (1:500 dilution; Sigma-Aldrich), and, for nestin/glial fibrillary acidic protein co-staining, mouse anti-rat nestin (1:1,000 dilution; BD Biosciences) and guinea-pig anti-glial fibrillary acidic protein (1:1,000 dilution; Advanced ImmunoChemical, Long Beach, CA) were used. Subsequently, cells were rinsed thoroughly and stained with secondary fluorescent-conjugated antibodies, including goat anti-rabbit Alexa 546 (1:250 dilution; Molecular Probes, Carlsbad, CA), goat anti-mouse Cy5 (1:250 dilution; Jackson ImmunoResearch, West Grove, PA), goat anti-mouse Cy3 (1:250 dilution; Jackson ImmunoResearch), or goat anti-guinea-pig Cy5 (1:250 dilution; Jackson ImmunoResearch). Stained cells were imaged by

fluorescence confocal microscopy (Leica Microsystems, Bannockburn, IL) or fluorescence microscopy (Nikon ECLIPSE TE2000-E, Nikon, Melville, NY, or Olympus IX50, Olympus America). When amplifying GFP signals, rabbit anti-GFP (1:2,000 dilution; Invitrogen, Center Valley, PA) and goat anti-rabbit Alexa 488 (1:250 dilution; Invitrogen) were used as primary and secondary antibodies, respectively.

Generation of mutant GFP NSC line. An internal ribosome entry site and puromycin resistance gene cassette was cloned into the *Eco* RI and *Xho* I sites of pFUGW (a kind gift from David Baltimore, California Institute of Technology, Pasadena, CA), which contains an enhanced GFP gene under the control of human ubiquitin-C promoter.⁴⁵ To insert an amber stop codon (TAA) into the GFP sequence, the substitution of "C" to "A" was introduced 603 nucleotides from the 5' end of GFP via QuikChange site-directed mutagenesis. The inserted point mutation was verified by DNA sequencing analysis.

A stable NSC cell population expressing the amber mutant GFP (NSC-mutGFP1) was generated by infection with lentiviral vector carrying the defective GFP. Briefly, 10 µg of pFUG carrying UB-mutGFP1-internal ribosome entry site and puromycin resistance gene cassette, 5 µg of pMDL g/p PRE, 3.5 µg of pcDNA3 IVS vesicular stomatitis virus G, and 1.5 µg of pRSV Rev, were transfected into HEK 293T cell lines (>70% confluency) by the calcium/phosphate transient transfection method.⁴⁶ The lentiviral vector was harvested and concentrated by ultracentrifugation (L8-55M Ultracentrifuge; Beckman-Coulter, Brea, CA), followed by resuspension in 100 µl of phosphate-buffered solution with 20% sucrose. Infectious titer, determined as previously described,⁴⁷ was $1.58 \times 10^7 \pm 4.66 \times 10^6$ IU/ml. NSCs were seeded onto 6-well tissue culture plates at a density of 3×10^5 cells/well and infected by 0, 1, 3, 5, 10, or 20 µl of the lentiviral vector. After puromycin selection,⁴⁸ the well in which 70–80% of cells survived was expanded for further analysis. The integration of lentiviral genomes was verified by cellular genomic DNA extraction, PCR amplification, and sequencing analysis of the PCR products. Importantly, GFP fluorescence was not detected in the mutGFP1 NSC cells, as confirmed by both flow cytometry and immunostaining against GFP.

Gene targeting assay. A point mutation (G to T) was inserted 16 nucleotides from the 5' end of GFP to yield a defective GFP open reading frame (mutGFP2). Subsequently, pFUG was digested with *Bam* HI and *Xho* I, and the mutGFP2-internal ribosome entry site and puromycin resistance gene cassette was subcloned into an AAV vector plasmid (*i.e.*, pAAV CMV GFP SN), where CMV GFP was excised by *Bam* HI and *Xho* I, such that the resulting targeting vector lacked a promoter. This genetic construct was packaged into the AAV2, AAV5, and AAV r3.45 capsids, and AAV r3.45 carrying luciferase driven by CMV promoter was used as a negative control. All viral vectors were harvested and purified as described earlier.

NSCs were seeded onto 24-well tissue culture plates at a density of 50,000 cells/well 1 day prior to AAV infection. Each AAV targeting vector was added to the NSCs at a genomic MOI of 5×10^5 , and the medium was replaced 24 hours later. Infected NSCs were analyzed by flow cytometry (Cytomics FC500; Beckman-Coulter) 3 days after infection, and the percentage of GFP⁺ cells was scored as the gene targeting frequency.

Cell sorting and sequencing. Cells continuously cultured for 14 days after AAV r3.45 infection were sorted with a Cytocpia Influx Sorter (UC Berkeley Cancer Center, Valencia, CA) to isolate FL1 (*i.e.*, GFP) positive cells. A fraction of the sorted cells was used for DNA sequencing analysis, and the remainder was expanded for an additional 26 days, at which time 94.5% still expressed GFP (data not shown). For DNA sequencing analysis, cellular genomic DNA was extracted with the QIAamp DNA Micro Kit (Qiagen) and amplified by using 5'-GCGGCAAGAACCCAAGTCT-3' and 5'-CGGCTTCGGCCAGTAACGTT-3' primers. To specifically amplify a fragment from the lentiviral vector and exclude amplification from the AAV viral genome, the forward primer was designed to amplify

from the half part of UB promoter in the targeted site, which existed only in the lentiviral vectors. The resulting PCR products were cloned using the StrataClone PCR cloning kit (Stratagene, La Jolla, CA), and individual clones were sequenced to confirm the gene correction.

SUPPLEMENTARY MATERIAL

Figure S1. Percentages of GFP positive neural stem cells transduced by recombinant vectors pseudotyped with capsids from AAV2, AAV5, and selected AAV variants.

Figure S2. Comparison of transduction efficiencies of selected AAV variants with wtAAV2 on a panel of cell types.

Figure S3. Comparison of transduction efficiencies of AAV r3.45 versus AAV2.

Figure S4. Comparison of transduction efficiencies of original AAV r3.45 (peptide insertion + V719M point mutation) versus modified AAV r3.45 (peptide insertion only, V719M correction).

Figure S5. Heparin column chromatograms for wild-type AAV2 and AAV r3.45 to investigate the heparin affinity of AAV r3.45.

Table S1. Genomic and infectious titers of wild-type AAV serotypes and AAV variants. Infectious titers were obtained from HEK 293T cell infection.

Table S2. Genomic titers of gene targeting vectors.

ACKNOWLEDGMENTS

This work was supported by California Institute for Regenerative Medicine Training grant number T1-00007 and grant number RT1-01021, and by National Research Foundation (NRF) grant funded by the Korea government (MEST) through the Active Polymer Center for Pattern Integration (No. R11-2007-050-00000-0). KIP was supported by Stem Cell Research Center grant and Healthcare Technology R&D Project grant funded by Korean Government. The contents of this publication are solely the responsibility of the authors and do not necessarily represent the official views of CIRM or any other agency of the State of California. The authors are also grateful to Prof Andrew Wurmser for kindly providing murine NSCs and Ilsun Kim for technical support on human NSC culture.

REFERENCES

- Hendrie, PC and Russell, DW (2005). Gene targeting with viral vectors. *Mol Ther* **12**: 9–17.
- Yates, F and Daley, GQ (2006). Progress and prospects: gene transfer into embryonic stem cells. *Gene Ther* **13**: 1431–1439.
- Madeira, C, Mendes, RD, Ribeiro, SC, Bours, JS, Aires-Barros, MR, da Silva, CL *et al.* (2010). Nonviral gene delivery to mesenchymal stem cells using cationic liposomes for gene and cell therapy. *J Biomed Biotechnol* **2010**: 735349.
- Jo, J and Tabata, Y (2008). Non-viral gene transfection technologies for genetic engineering of stem cells. *Eur J Pharm Biopharm* **68**: 90–104.
- Singer, O, Tiscornia, G, Ikawa, M and Verma, IM (2006). Rapid generation of knockdown transgenic mice by silencing lentiviral vectors. *Nat Protoc* **1**: 286–292.
- Marumoto, T, Tashiro, A, Friedmann-Morvinski, D, Scadeng, M, Soda, Y, Gage, FH *et al.* (2009). Development of a novel mouse glioma model using lentiviral vectors. *Nat Med* **15**: 110–116.
- Fan, X, Valdimarsdottir, G, Larsson, J, Brun, A, Magnusson, M, Jacobsen, SE *et al.* (2002). Transient disruption of autocrine TGF-beta signaling leads to enhanced survival and proliferation potential in single primitive human hemopoietic progenitor cells. *J Immunol* **168**: 755–762.
- Silva, J, Barrandon, O, Nichols, J, Kawaguchi, J, Theunissen, TW and Smith, A (2008). Promotion of reprogramming to ground state pluripotency by signal inhibition. *PLoS Biol* **6**: e253.
- Hester, ME, Song, S, Miranda, CJ, Eagle, A, Schwartz, PH and Kaspar, BK (2009). Two factor reprogramming of human neural stem cells into pluripotency. *PLoS ONE* **4**: e7044.
- Gropp, M, Itsykson, P, Singer, O, Ben-Hur, T, Reinhartz, E, Galun, E *et al.* (2003). Stable genetic modification of human embryonic stem cells by lentiviral vectors. *Mol Ther* **7**: 281–287.
- Suzuki, K, Mitsui, K, Aizawa, E, Hasegawa, K, Kawase, E, Yamagishi, T *et al.* (2008). Highly efficient transient gene expression and gene targeting in primate embryonic stem cells with helper-dependent adenoviral vectors. *Proc Natl Acad Sci USA* **105**: 13781–13786.
- Thomas, CE, Ehrhardt, A and Kay, MA (2003). Progress and problems with the use of viral vectors for gene therapy. *Nat Rev Genet* **4**: 346–358.
- Conrad, C, Gupta, R, Mohan, H, Niess, H, Bruns, CJ, Kopp, R *et al.* (2007). Genetically engineered stem cells for therapeutic gene delivery. *Curr Gene Ther* **7**: 249–260.
- Lim, KI, Klimczak, R, Yu, JH and Schaffer, DV (2010). Specific insertions of zinc finger domains into Gag-Pol yield engineered retroviral vectors with selective integration properties. *Proc Natl Acad Sci USA* **107**: 12475–12480.

15. Urnov, FD, Miller, JC, Lee, YL, Beausejour, CM, Rock, JM, Augustus, S *et al.* (2005). Highly efficient endogenous human gene correction using designed zinc-finger nucleases. *Nature* **435**: 646–651.
16. Verma, IM and Weitzman, MD (2005). Gene therapy: twenty-first century medicine. *Annu Rev Biochem* **74**: 711–738.
17. Bennicelli, J, Wright, JF, Komaromy, A, Jacobs, JB, Hauck, B, Zelenia, O *et al.* (2008). Reversal of blindness in animal models of leber congenital amaurosis using optimized AAV2-mediated gene transfer. *Mol Ther* **16**: 458–465.
18. Maheshri, N, Koerber, JT, Kaspar, BK and Schaffer, DV (2006). Directed evolution of adeno-associated virus yields enhanced gene delivery vectors. *Nat Biotechnol* **24**: 198–204.
19. Kaplitt, MG, Leone, P, Samulski, RJ, Xiao, X, Pfaff, DW, O'Malley, KL *et al.* (1994). Long-term gene expression and phenotypic correction using adeno-associated virus vectors in the mammalian brain. *Nat Genet* **8**: 148–154.
20. Flannery, JC, Zolotukhin, S, Vaquero, MI, LaVail, MM, Muzyczka, N and Hauswirth, WW (1997). Efficient photoreceptor-targeted gene expression *in vivo* by recombinant adeno-associated virus. *Proc Natl Acad Sci USA* **94**: 6916–6921.
21. Russell, DW and Hirata, RK (1998). Human gene targeting by viral vectors. *Nat Genet* **18**: 325–330.
22. Chamberlain, JR, Schwarze, U, Wang, PR, Hirata, RK, Hankenson, KD, Pace, JM *et al.* (2004). Gene targeting in stem cells from individuals with osteogenesis imperfecta. *Science* **303**: 1198–1201.
23. Rogers, CS, Hao, Y, Rokhlina, T, Samuel, M, Stoltz, DA, Li, Y *et al.* (2008). Production of CFTR-null and CFTR-DeltaF508 heterozygous pigs by adeno-associated virus-mediated gene targeting and somatic cell nuclear transfer. *J Clin Invest* **118**: 1571–1577.
24. Smith-Arica, JR, Thomson, AJ, Ansell, R, Chiorini, J, Davidson, B and McWhir, J (2003). Infection efficiency of human and mouse embryonic stem cells using adenoviral and adeno-associated viral vectors. *Cloning Stem Cells* **5**: 51–62.
25. Hughes, SM, Moussavi-Harami, F, Sauter, SL and Davidson, BL (2002). Viral-mediated gene transfer to mouse primary neural progenitor cells. *Mol Ther* **5**: 16–24.
26. Temple, S (2001). The development of neural stem cells. *Nature* **414**: 112–117.
27. Lai, K, Kaspar, BK, Gage, FH and Schaffer, DV (2003). Sonic hedgehog regulates adult neural progenitor proliferation *in vitro* and *in vivo*. *Nat Neurosci* **6**: 21–27.
28. Koerber, JT, Klimczak, R, Jang, JH, Dalkara, D, Flannery, JG and Schaffer, DV (2009). Molecular evolution of adeno-associated virus for enhanced glial gene delivery. *Mol Ther* **17**: 2088–2095.
29. Koerber, JT, Jang, JH and Schaffer, DV (2008). DNA shuffling of adeno-associated virus yields functionally diverse viral progeny. *Mol Ther* **16**: 1703–1709.
30. Excoffon, KJ, Koerber, JT, Dickey, DD, Murtha, M, Keshavjee, S, Kaspar, BK *et al.* (2009). Directed evolution of adeno-associated virus to an infectious respiratory virus. *Proc Natl Acad Sci USA* **106**: 3865–3870.
31. Klimczak, RR, Koerber, JT, Dalkara, D, Flannery, JG and Schaffer, DV (2009). A novel adeno-associated viral variant for efficient and selective intravitreal transduction of rat Müller cells. *PLoS ONE* **4**: e7467.
32. Müller, OJ, Kaul, F, Weitzman, MD, Pasqualini, R, Arap, W, Kleinschmidt, JA *et al.* (2003). Random peptide libraries displayed on adeno-associated virus to select for targeted gene therapy vectors. *Nat Biotechnol* **21**: 1040–1046.
33. Opie, SR, Warrington, KH Jr, Agbandje-McKenna, M, Zolotukhin, S and Muzyczka, N (2003). Identification of amino acid residues in the capsid proteins of adeno-associated virus type 2 that contribute to heparan sulfate proteoglycan binding. *J Virol* **77**: 6995–7006.
34. Hansen, J, Qing, K, Kwon, HJ, Mah, C and Srivastava, A (2000). Impaired intracellular trafficking of adeno-associated virus type 2 vectors limits efficient transduction of murine fibroblasts. *J Virol* **74**: 992–996.
35. Ferrari, FK, Samulski, T, Shenk, T and Samulski, RJ (1996). Second-strand synthesis is a rate-limiting step for efficient transduction by recombinant adeno-associated virus vectors. *J Virol* **70**: 3227–3234.
36. Kim, JB, Greber, B, Araújo-Bravo, MJ, Meyer, J, Park, KI, Zaehres, H *et al.* (2009). Direct reprogramming of human neural stem cells by OCT4. *Nature* **461**: 649–643.
37. Kim, JB, Sebastiano, V, Wu, G, Araújo-Bravo, MJ, Sasse, P, Gentile, L *et al.* (2009). Oct4-induced pluripotency in adult neural stem cells. *Cell* **136**: 411–419.
38. Palmer, TD, Markakis, EA, Willhoite, AR, Safar, F and Gage, FH (1999). Fibroblast growth factor-2 activates a latent neurogenic program in neural stem cells from diverse regions of the adult CNS. *J Neurosci* **19**: 8487–8497.
39. Kim, HT, Kim, IS, Lee, IS, Lee, JP, Snyder, EY and Park, KI (2006). Human neurospheres derived from the fetal central nervous system are regionally and temporally specified but are not committed. *Exp Neurol* **199**: 222–235.
40. Perabo, L, Büning, H, Kofler, DM, Ried, MU, Girod, A, Wendtner, CM *et al.* (2003). *In vitro* selection of viral vectors with modified tropism: the adeno-associated virus display. *Mol Ther* **8**: 151–157.
41. Stemmer, WP (1994). Rapid evolution of a protein *in vitro* by DNA shuffling. *Nature* **370**: 389–391.
42. Zhao, H and Arnold, FH (1997). Optimization of DNA shuffling for high fidelity recombination. *Nucleic Acids Res* **25**: 1307–1308.
43. Batard, P, Jordan, M and Wurm, F (2001). Transfer of high copy number plasmid into mammalian cells by calcium phosphate transfection. *Gene* **270**: 61–68.
44. Koerber, JT, Maheshri, N, Kaspar, BK and Schaffer, DV (2006). Construction of diverse adeno-associated viral libraries for directed evolution of enhanced gene delivery vehicles. *Nat Protoc* **1**: 701–706.
45. Lois, C, Hong, EJ, Pease, S, Brown, EJ and Baltimore, D (2002). Germine transmission and tissue-specific expression of transgenes delivered by lentiviral vectors. *Science* **295**: 868–872.
46. Yu, JH and Schaffer, DV (2006). Selection of novel vesicular stomatitis virus glycoprotein variants from a peptide insertion library for enhanced purification of retroviral and lentiviral vectors. *J Virol* **80**: 3285–3292.
47. Leonard, JN, Shah, PS, Burnett, JC and Schaffer, DV (2008). HIV evades RNA interference directed at TAR by an indirect compensatory mechanism. *Cell Host Microbe* **4**: 484–494.
48. Peltier, J, O'Neill, A and Schaffer, DV (2007). PI3K/Akt and CREB regulate adult neural hippocampal progenitor proliferation and differentiation. *Dev Neurobiol* **67**: 1348–1361.

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US17/32542

A. CLASSIFICATION OF SUBJECT MATTER

IPC - C07K 14/075; C12Q 1/70; C12N 15/62, 15/63, 15/861 (2017.01)

CPC - C07K 14/075; C12Q 1/70; C12N 15/62, 15/63, 15/861

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

See Search History document

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

See Search History document

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

See Search History document

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	US 2015/0232953 A1 (THE REGENTS OF THE UNIVERSITY OF CALIFORNIA) 20 August, 2015; paragraphs [0071], [0085]-[0087], [0094]	1-13
A	(MONDEGO, JM et al.) A Genome Survey of Moniliophthora perniciosa gives New Insights into Witches' Broom Disease of Cacao. BMC Genomics. 18 November, 2008; Vol. 18, No. 9; pages 1-25; page 6, column 2, paragraph 3; Genbank supplement pages 1-2; DOI: 10.1186/1471-2164-9-548	1-13
A	(ASURI, P et al.) Directed Evolution of Adeno-associated Virus for Enhanced Gene Delivery and Gene Targeting in Human Pluripotent Stem Cells. Molecular Therapy. 22 November, 2011; Vol. 20, No. 2; pages 329-338; page 332, column 1, paragraph 1; DOI: 10.1038/mt.2011.255	1-13
A	(XIAO, W et al.) Gene Therapy Vectors Based on Adeno-Associated Virus Type 1. Journal of Virology. May, 1999; Vol. 73, No. 5; pages 3994-4003; page 3998, column 2, paragraph 1; Genbank supplement pages 1-2	26-43
A	WO 2015/164757 A1 (OREGON HEALTH AND SCIENCE UNIVERSITY) 29 October, 2015; pages 25-26	26-43
A	(JANG, JH et al.) An Evolved Adeno-Associated Viral Variant Enhances Gene Delivery and Gene Targeting in Neural Stem Cells. Molecular Therapeutics. 11 January, 2011; Vol. 19, No. 4; pages 667-675; page 669, column 1, paragraph 2; DOI: 10.1038/mt.2010.287	26-43



Further documents are listed in the continuation of Box C.



See patent family annex.

* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search

04 October 2017 (04.10.2017)

Date of mailing of the international search report

24 OCT 2017

Name and mailing address of the ISA/

Mail Stop PCT, Attn: ISA/US, Commissioner for Patents

P.O. Box 1450, Alexandria, Virginia 22313-1450

Facsimile No. 571-273-8300

Authorized officer

Shane Thomas

PCT Helpdesk: 571-272-4300

PCT OSP: 571-272-7774

INFORME DE BÚSQUEDA INTERNACIONAL		Nro. de solicitud internacional PCT/US17/32542
A. CLASIFICACION DEL OBJETO DE LA SOLICITUD IPC: C07K 14/075; C12Q 1/70; C12N 15/62, 15/63, 15/861 (01.2017) CPC: C07K 14/075; C12Q 1/70; C12N 15/62, 15/63, 15/861 Según la Clasificación Internacional de Patentes o, ambas, la clasificación nacional y la CIP		
B. CAMPO DE BÚSQUEDA		
Documentación mínima buscada (sistema de clasificación seguido de símbolo de clasificación) Ver historial de búsqueda		
Otros documentos buscados distintos de la documentación mínima que hayan sido incluidos en el campo de búsqueda Ver historial de búsqueda		
Base de datos electrónica consultada durante la búsqueda internacional (nombre de la base de datos y, si fuera posible, términos utilizados en la búsqueda) Ver historial de búsqueda		
C. DOCUMENTOS CONSIDERADOS RELEVANTES		
Categoría*	Cita del documento con indicación de los párrafos relevantes si correspondiese	Relevante para la reivindicación No.
A	US 2015/0232953 A1 (THE REGENTS OF THE UNIVERSITY OF CALIFORNIA) 20 August, 2015; paragraphs [0071j, [0085]-[0087], [0094]	1-13
A	(MONDEGO, JM et al.) A Genome Survey of Moniliophthora perniciosa gives New Insights into Witches' Broom Disease of Cacao. BMC Genomics. 18 November, 2008; Vol. 18, No. 9; pages 1-25; page 6, column 2, paragraph 3; Genbank supplement pages 1-2; DOI:10.1186/1471-2164-9-548	1-13
A	(ASURI, P et at.) Directed Evolution of Adeno-associated Virus for Enhanced Gene Delivery and Gene Targeting in Human Pluripotent Stem Cells. Molecular Therapy. 22 November, 2011; Vol. 20, No. 2; pages 329-338; page 332, column 1, paragraph 1; DOI: 10.1038/mt.2011.255	1-13
A	(XIAO, W et al.) Gene Therapy Vectors Based on Adeno-Associated Virus Type 1. Journal of Virology. May, 1999; Vol. 73, No. 5; pages 3994-4003; page 3998, column 2, paragraph 1; Genbank supplement pages 1-2	26-43
A	WO 2015/164757 A1 (OREGON HEALTH AND SCIENCE UNIVERSITY) 29 October, 2015; pages 25-26	26-43

A	(JANG, JH et al.) An Evolved Adeno-Associated Viral Variant Enhances Gene Delivery and Gene Targeting in Neural Stem Cells. Molecular Therapeutics. 11 January, 2011; Vol. 19, No. 4; pages 667-675; page 669, column 1, paragraph 2; DOI: 10.1038/mt.2010.287	26-43
☐ En la continuación del casillero C figuran otros documentos ☐ Ver anexo de familia de patentes		
*Categorías especiales de los documentos citados "A" documento que define el estado general de la técnica no considerado como particularmente relevante. "E" solicitud de patente o patente anterior pero publicada en la fecha de presentación internacional o en fecha posterior. "L" documento que puede plantear dudas sobre una reivindicación de prioridad o que se cita para determinar la fecha de publicación de otra cita o por otra razón especial (según se indique). "O" documento que se refiere a una divulgación oral, a una utilización, a una exposición o a cualquier otro medio. "P" documento publicado antes de la fecha de presentación internacional, pero con posterioridad a la fecha de prioridad reivindicada.		"T" documento posterior publicado con posterioridad a la fecha de presentación internacional o de prioridad que no pertenece al estado de la técnica pertinente pero que se cita por permitir la comprensión del principio o teoría que constituye la base de la invención. "X" documento particularmente relevante; la invención reivindicada no puede considerarse nueva o que implique una actividad inventiva por referencia al documento aisladamente considerado. "Y" documento particularmente relevante; la invención reivindicada no puede considerarse que implique una actividad inventiva cuando el documento se asocia a otro u otros documentos de la misma naturaleza, cuya combinación resulta evidente para un experto en la materia. "&" documento que forma parte de la misma familia de patentes
Fecha en la que efectivamente se concluyó la búsqueda internacional 4 de octubre de 2017		Fecha de envío del informe de búsqueda internacional 24 de octubre 2017
Nombre y dirección de correo de la ABI Mail Stop PCT, Attn: ISA/US, Commissioner for Patents, P.O. Box 1450, Alexandria, Virginia 22313-1450, Fax No. 571-273-8300		Funcionario autorizado Shane Thomas Mesa de ayuda del PCT: 571-272-4300 PCT OSP: 571-272-7774

Formulario PCT/ISA/210 (Segunda hoja) (enero de 2015)