Supporting Information

Gladyshev and Arkhipova 10.1073/pnas.1100266108

SI Materials and Methods

Library Manipulations. The *A. vaga* (6× coverage) and *P. roseola* (4× coverage) genomic libraries are described in ref. 1. For library screening, we used ^{32}P-labeled PCR fragments spanning the central region of RVT from *A. vaga* and *P. roseola*. Hybridizing fosmids were selected, sheared to ∼2-kb fragments by sonication, blunt-ended with T4 DNA polymerase, subcloned into pBluescript II SK−, and sequenced on the ABI3730XL at the W. M. Keck Ecological and Evolutionary Genetics Facility at the Marine Biological Laboratory. Subclones were assembled into contigs with Phrap/Phred/Consed (CodonCode). Any remaining gaps were closed by primer walking. Sequences reported in this study have been deposited in the GenBank database (accession nos. JN235987–JN235989).

Plasmid construction

Plasmid	Description
pEAG5.1	Full-length NcRVT was amplified from FGSC 2489 with primers ATGGCCGCCTCTTCAGAA and AATAATGAATTCTTAGCCCTGCCAATGGAAC and inserted into pBlueScriptII SK+ digested with HincII and EcoRI.
pEAG5.2	Plasmid pEAG5.1 was amplified with primers GCCGACATGTGGCTTTG and GTGAAGCCGGTAGAGAAG and circularized, yielding a replacement of the YRLHDD motif with YRLHAD.
pEAG19	A 6xHis epitope (annealed primers CTAGAATGGCACATCACCACCACCATCACGTGGGTTTAAT and TAAACCCACGTGATGGTGGTGGTGATGTGCCATT) was inserted into pMF272 (2) as a XbaI/PacI fragment.
pEAG20.1	Full-length wild-type NcRVT was amplified from pEAG5.1 with primers AATAATTTAATTAACATGGCCGCCTCTTCAGAA and AATAATGAATTCTTAGCCCTGCCAATGGAACC and inserted into pEAG19 as a PacI/XbaI fragment.
pEAG20.2	Full-length mutant NcRVT (D529A) was amplified from pEAG5.2 with primers AATAATTTAATTAACATGGCCGCCTCTTCAGAA and AATAATGAATTCTTAGCCCTGCCAATGGAACC and inserted into pEAG19 as a PacI/XbaI fragment.

Strains

Strain ID	Genotype	Notes
FGSC 2489	74-OR23-1VA	Wild type
FGSC 2225	Mauriceville-1c	Wild type
FGSC 6103	*his-3; matA*	*his-3* allele 1-234-723
FGSC 4264	*cpc-1; matA*	*cpc-1* allele CD-15
dRVT1	*rvt::hygR+; mat A*	KO strain
dRVT2	*rvt::hygR+; mat a*	KO strain
G1004	*his-3; rvt::hygR+; mat A*	This study
G0021	*his-3+::6xHis-RVT(D529A); rvt::hygR+; mat A*	This study
G0022	*his-3+::6xHis-RVT(WT); rvt::hygR+; mat A*	This study

Plasmids pEAG20.1 and pEAG20.2 were linearized with NdeI and transformed into G1004 by electroporation; primary transformants were screened by PCR and Southern blots, and homokaryons were purified by microconidiation to generate strains G0021 (D529A) and G0022 (wild type NcRVT). *N. crassa* wildtype and mutant strains, including knockouts for NcRVT (NCU09536), were obtained from the Fungal Genetics Stock Center (FGSC) (17, 18).

RT-PCR. Ten- to 15-d-old macroconidia were inoculated into 1× Vogel's medium containing 1.5% sucrose, and grown overnight at 30 °C and constant light while shaking at 180 rpm. Overexpression of endogenous NcRVT was induced for 3–4 h, and 23–25 mg of squeeze-dried mycelium was thoroughly ground with ∼90 mg (100 μL) of washed sea sand (FisherChemical) and 1 mL of TRIzol reagent (Invitrogen). Total RNA was extracted following manufacturer's instructions and resuspended in 100 μL of water. Three microliters of RNA was used for RT with SuperScript II (Invitrogen) and 100 pmol of N10 primer in the final volume of 20 μL, following manufacturer's instructions. RT reactions were diluted 1:20, and 1 μL of each dilution was used for PCR in the final volume of 12 μL. The following primer combinations were used for semiquantitative RT-PCR:

Marker	Forward	Reverse	Size, bp
RVT1 (central region)	CTGGATGAAAA GAACCCCCTCT	TGAACTCCGTCA CTACCTTCCA	604
RVT2 (splice junction)	GAGCTCAATCGT CACCATCAAT	CGTACATGCTGG CAAATCTGT	438
NCU01640	GGCCACTCTTCA TCCTAGCTTT	GGGTGTCATGTC CTGATTGTGT	613
NCU08641	CCAGGAGTCCCT TCAACATTTC	GAGCGTAACCTT GCATTTCCTT	497
NCU09802	CGGCTCAGCAA TACTACCAACA	CGGCCATCTATC TCGTGTACTA	455
NCU05498	CATGAACATCAA CGACCTGGAG	GGCAGCCTTCTT CTTCTTCTTG	400
NCU06110	CTGTTCACTTCGA CTGTGCTGA	ACTTGTCGTTCTG AGCCTTGC	550
pMauriceville	TGTGCCCCTTCTA CGTTACAGA	GAGAGCTTGCGT TGAACAAGTC	605

The forward primer of RVT2 can anneal only to cDNA corresponding to spliced mRNA, eliminating the possibility of unwanted amplification of genomic DNA. Markers NCU05498 and NCU06110 were selected on the basis of their high expression levels that remained unaffected by low doses of blasticidin and cycloheximide used in this study.

Vegetative Growth Assays. The degree of vegetative growth inhibition was assayed by placing 2 μL of macroconidial suspension at the center of a 15-cm Petri dish containing 1× Vogel's agar supplemented with 0.1 μg/mL of blasticidin or cycloheximide and allowing conidia to germinate and grow for 18 h at 30 °C. Mycelial boundary was outlined with a marker and plates were incubated for additional 12 h, after which the second mycelial outline was drawn, and the distance between mycelial fronts at 18 and 30 h was measured and recorded in eight random di-

rections per plate. Three plates were scored per each strain/antibiotic combination.

Protein Purification. NcRVT was purified from strains G0021, G0022, and FGSC 2225. Overexpression of endogenous NcRVT was induced in FGSC 2225 by supplementing a 12-h-old culture with 0.1 μg/mL of blasticidin and growing it for additional 12 h, after which mycelial pads were harvested. Purification schemes were identical for all strains and included the following three steps: (*a*) preparation of mycelial lysates, (*b*) ultracentrifugation in a sucrose gradient, and (*c*) small-scale ion-exchange chromatography using DEAE Sepharose CL-6B (Pharmacia).

Lysates were prepared by two alternative procedures starting from 24-h-old squeeze-dried mycelial mats. Lysis procedure 1: 1–1.5 g of squeeze-dried mycelium was thoroughly ground with washed sea sand in 5 mL of 0.5 M ammonium sulfate (pH 7.5) at 4 °C and centrifuged briefly at 12000 × *g* to pellet debris. Solid ammonium sulfate was added to the supernatant at the concentration of 187 g/L, tubes were swirled for 15 min at 4 °C, and proteins were precipitated by centrifugation for 10 min at 9,000 × *g* and discarded. RVT-containing fraction was precipitated by adding 82 g/L of ammonium sulfate, swirling for 15 min, and centrifuging for 10 min at 9,000 × *g*. The pellet was dissolved in 1 mL of buffer A (100 mM NaCl, 50 mM Tris·HCl, pH 7.5, 1:1,000 β-mercaptoethanol) containing protein inhibitors (Halt Protease FGSC Inhibitor Mixture; Pierce). Lysis procedure 2: 1–1.5 g of squeeze-dried mycelium was thoroughly ground in liquid nitrogen and mixed with 3 mL of buffer A containing protein inhibitors (Halt Protease FGSC Inhibitor Mixture; Pierce). The mixture was slowly thawed at 4 °C, vortexed, and centrifuged for 10 min at 12,000 × *g* to pellet debris.

Lysates were loaded on sucrose density gradients made from 17 mL of 40% sucrose and 19.6 mL of 20% sucrose, with both stock solutions prepared in buffer A. Gradients were centrifuged in a SW28 rotor (Beckman) for 30 h at 25,000 rpm, and 1-mL fractions were taken with a peristaltic pump, starting from the bottom of each tube. Fractions were loaded on an SDS/PAGE gel and analyzed by staining with GelCode Blue (Pierce) or Western blots, using His-tag monoclonal antibody (mouse IgG1; Novagen 70796-3, Lot N54326).

DEAE ion exchange chromatography was carried out in a 5-mL gravity column packed with 0.4 mL of Fast Flow DEAE Sepharose CL-6B (Pharmacia). The column was preequilibrated with buffer A at 4 °C, and two to three sucrose fractions containing NcRVT were passed through the column. The column was washed with 5 vol of buffer A, and NcRVT was eluted either with buffer B (200 mM NaCl, 50 mM Tris·HCl, pH 7.5, 1:1,000 β-mercaptoethanol) or buffer A containing 5 mM $MgCl_2$. Whereas several copurifying proteins were detected by GelCode Blue staining in the buffer B eluate, elution with buffer A + $MgCl_2$ yielded nearly pure NcRVT protein.

Terminal Transferase Activity Assays. A master mix for 10 reactions was prepared by combining 20 μL of DEAE Sepharose eluate (buffer A + $MgCl_2$) containing 0.01–0.05 mg/mL (6xHis-) or 0.1–0.3 mg/mL (induced) NcRVT with 40 μL of 2× buffer A (200 mM NaCl, 100 mM Tris·HCl, pH 7.5), 18 μL of water, and 2 μL of [α-^{32}P]dATP (or [α-^{32}P]CTP). Eight microliters of this master mix was aliquoted into each reaction tube, supplemented with 3 mM $MnCl_2$ to the final volume of 9 μL, and reactions were allowed to proceed at room temperature for 1 h. Reactions were then supplemented with 1 μL of 10 mM (d)NTP and incubated for another 1 h, after which 1 μL of 20 mg/mL Proteinase K was added. Without any further purification 3 μL of each digested reaction was mixed with 7 μL of loading buffer and run on a 12% denaturing PAGE.

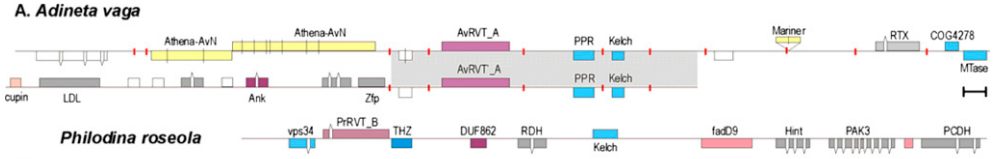
Determination of Molecular Weight. The approximate molecular weight of the NcRVT complex was determined by comparing its sedimentation velocity to those of the three reference protein complexes—thyroglobulin (660 kDa; Sigma), catalase (250 kDa; Sigma), and aldolase (160 kDa; Sigma)—dissolved in buffer A at 2 mg/mL and centrifuged in a sucrose gradient with parameters identical to those used during the initial protein purification. Each gradient contained thyroglobulin as a reference and purified 6xHis-RVT or catalase or aldolase. Fractions (0.5 mL) were taken to determine the position of NcRVT relative to thyroglobulin. Marker positions were determined by measuring absorbance at 280 nm of each fraction, and NcRVT position was determined by Western blot analysis. The exact position of any given peak was determined by the weighted average formula: $(Fl \times (Al - An) + Fm \times (Am - An))/((Al - An) + (Am - An))$, where Fm and Fl are the adjacent fraction numbers containing the highest and the second highest amount of the protein, and Am, Al, and An are the actual amounts of the protein found in the three adjacent fractions, where Am > Al > An.

Size Exclusion Chromatography. The peak sucrose gradient fraction containing NcRVT was loaded at 0.8 mL/min onto a gel filtration column (Superose 6 10/300 GL; GE Healthcare), preequilibrated with buffer A. The column was eluted with the same buffer at 0.8 mL/min. Fractions were loaded on an SDS/PAGE gel and analyzed by staining with GelCode Blue (Pierce) or Western blots, using His-tag monoclonal antibody (mouse IgG1; Novagen 70796-3, Lot N54326).

Bioinformatics. Synteny in genomic contigs up to 100 kb in length was examined with the aid of the ACT comparison tool (www.sanger.ac.uk/resources/software/act/) at the European Bioinformatics Institute website. Synonymous (d_S) and nonsynonymous (d_N) substitution rates per site were calculated by CODEML in the PAML package (3). Structure-based alignments were generated on the PROMALS3D server (4). Alignment of *rvt* and other RT-related sequences was done by submitting the aligned dataset (5) to the HHpred server (6) for profile–profile comparison and readjustment to include scoring for the secondary structure. Predicted secondary structures were validated on the Jpred server (7) and visualized with STRAP (http://www.bioinformatics.org/strap) with minor manual adjustments. Phylogenetic analysis by the minimum evolution method was done with MEGA 5.1 (8), using parameters estimated by ProtTest (9), and by maximum likelihood with RAxML version 7.2.8 (10), using the WAG substitution matrix. Evaluation of the phylogenetic data structure using phylogenetic networks was done with NeighborNet (11), implemented in SplitsTree 4.10 (12). Likelihood distance-based phylogenetic trees were inferred by applying the BioNJ algorithm (13) in SplitsTree 4.10 on ProteinML distances computed by using the WAG model and the parameter values estimated by ProtTest. NeighborNet networks were constructed from the same distance estimates.

1. Hur JH, Van Doninck K, Mandigo ML, Meselson M (2009) Degenerate tetraploidy was established before bdelloid rotifer families diverged. *Mol Biol Evol* 26:375–383.
2. Freitag M, Hickey PC, Raju NB, Selker EU, Read ND (2004) GFP as a tool to analyze the organization, dynamics and function of nuclei and microtubules in *Neurospora crassa*.. *Fungal Genet Biol* 41:897–910.
3. Yang Z (1997) PAML: A program package for phylogenetic analysis by maximum likelihood. *Comput Appl Biosci* 13:555–556.
4. Pei J, Kim BH, Grishin NV (2008) PROMALS3D: A tool for multiple protein sequence and structure alignments. *Nucleic Acids Res* 36:2295–2300.
5. Arkhipova IR, Pyatkov KI, Meselson M, Evgen'ev MB (2003) Retroelements containing introns in diverse invertebrate taxa. *Nat Genet* 33:123–124.
6. Söding J (2005) Protein homology detection by HMM-HMM comparison. *Bioinformatics* 21:951–960.
7. Cole C, Barber JD, Barton GJ (2008) The Jpred 3 secondary structure prediction server. *Nucleic Acids Res* 36(Web Server issue, Suppl 2):W197–W201.

8. Tamura K, Dudley J, Nei M, Kumar S (2007) MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. *Mol Biol Evol* 24:1596–1599.
9. Abascal F, Zardoya R, Posada D (2005) ProtTest: Selection of best-fit models of protein evolution. *Bioinformatics* 21:2104–2105.
10. Stamatakis A, Hoover P, Rougemont J (2008) A rapid bootstrap algorithm for the RAxML Web servers. *Syst Biol* 57:758–771.
11. Bryant D, Moulton V (2004) Neighbor-net: An agglomerative method for the construction of phylogenetic networks. *Mol Biol Evol* 21:255–265.
12. Huson DH, Bryant D (2006) Application of phylogenetic networks in evolutionary studies. *Mol Biol Evol* 23:254–267.
13. Gascuel O (1997) BIONJ: An improved version of the NJ algorithm based on a simple model of sequence data. *Mol Biol Evol* 14:685–695.
14. Gladyshev EA, Meselson M, Arkhipova IR (2008) Massive horizontal gene transfer in bdelloid rotifers. *Science* 320:1210–1213.
15. Huang H, Chopra R, Verdine GL, Harrison SC (1998) Structure of a covalently trapped catalytic complex of HIV-1 reverse transcriptase: Implications for drug resistance. *Science* 282:1669–1675.
16. Mitchell M, Gillis A, Futahashi M, Fujiwara H, Skordalakes E (2010) Structural basis for telomerase catalytic subunit TERT binding to RNA template and telomeric DNA. *Nat Struct Mol Biol* 17:513–518.
17. Colot HV, Park G, Turner GE, Ringelberg C, Crew CM, Litvinkova L, Weiss RL, Borkovich KA, Dunlap JC (2006) A high-throughput gene knockout procedure for Neurospora reveals functions for multiple transcription factors. *Proc Natl Acad Sci USA* 103:10352–10357.
18. McCluskey K, Wiest A, Plamann M (2010) The Fungal Genetics Stock Center: A repository for 50 years of fungal genetics research. *J Biosci* 35:119–126.



B.

Gene ID, name	Contig ID	AI	% identity to best hit	Best hit, E-value	Best hit, metazoan	Best hit, taxonomy	Definition
PR002_RVT	128E9	414	39	1e-180	no hits	Eukaryota; Fungi	PrRVT-B
AV003_RVT	111A10	182	27	7e-80	no hits	Eukaryota; Fungi	AvRVT-A
AV021_RVT	116B20	181	27	1e-79	no hits	Eukaryota; Fungi	AvRVT-A'
PR007_fadD9	128E9	91	29	1e-106	4e-67	Eukaryota; Amoebozoa	fatty-acid-CoA ligase fadD9
PR006_Kelch	128E9	85	46	1e-66	2e-29	Bacteria	kelch domain-containing protein
PR003_THZ	128E9	84	47	8e-57	3e-20	(Archaea/Bacteria)	hydroxyethylthiazole kinase
PR011	128E9	82	34	2e-36	no hits	Eukaryota; Amoebozoa	hypothetical protein
AV005_Kelch	111A10	40	43	2e-31	6e-14	Bacteria	kelch domain-containing protein
AV009_MTase	111A10	34	31	1e-15	no hits	Bacteria	methyltransferase
AV023_Kelch	116B20	33	48	8e-26	2e-11	Bacteria	kelch domain-containing protein
PR001_ vps34	128E9	29	31	1e-13	no hits	(Eukaryota/Bacteria)	phosphatidylinositol 3-kinase vps34
AV004_PPR	111A10	16	27	8e-08	no hits	Bacteria	serine/threonine protein kinase, PPR
AV022_PPR	116B20	15	27	2e-07	no hits	Bacteria	serine/threonine kinase, PPR
AV012_cupin	116B20	14	33	8e-07	no hits	Eukaryota; Amoebozoa	cupin/spherulin-like protein
AV008_cog4278	111A10	10	37	2e-05	no hits	(Bacteria/Eukaryota)	COG4278: conserved protein
AV019_Zfp	116B20	-8	33	8e-08	8e-08	Eukaryota; Metazoa	zinc finger protein 821-like
PR009_Hint	128E9	-10	30	1e-04	1e-04	Eukaryota, Metazoa	hedgehog/Hint domain protein
AV007_RTX	111A10	-11	27	9e-09	9e-09	(Bacteria/Eukaryota)	COG2931: RTX toxins Ca2+-binding
PR005_RDH	128E9	-20	42	3e-50	3e-50	Eukaryota; Metazoa	retinol dehydrogenase
AV013_LDL	116B20	-39	29	2e-17	2e-17	Eukaryota; Metazoa	LDL receptor-related protein 1-like
PR010_PAK3	128E9	-101	53	1e-140	1e-140	Eukaryota, Metazoa	serine/threonine-protein kinase PAK3
PR012_PCDH	128E9	-130	28	5e-57	5e-57	Eukaryota; Metazoa	protocadherin alpha-C2 isoform 1

C.

Bdelloid Species	Lineage A	Lineage B
Adineta vaga **(Adinetidae)**	+	+
Habrotrocha rosa **(Habrotrochidae)**	+	n.d.
Philodina roseola **(Philodinidae)**	n.d.	+
Philodina acuticornis **(Philodinidae)**	+	+
Macrotrachela quadricornifera **(Philodinidae)**	+	+

Fig. S1. *rvt* genes in bdelloid rotifers. (*A*) Genomic environment of *rvt* in sequenced *A. vaga* and *P. roseola* fosmids. ORFs are colored according to their putative origin: metazoan, gray; bacterial, blue; fungal, purple; unknown hypothetical, white; transposable element, yellow. Intron positions are indicated by V-shaped lines and frameshifts and in-frame stop codons by vertical lines. Gray shading designates the region of colinearity between two members of the *A. vaga* allelic pair up to the breakpoint junction, which contains a stretch of telomeric repeats. Although there is a Kelch-repeat–containing protein on both *A. vaga* and *P. roseola* contigs, its presence cannot be regarded as evidence of synteny, because similar proteins are frequently found in other subtelomeric regions (14). (Scale bar, 1 kb.) (*B*) Functional characteristics of ORFs depicted in *A*, as inferred from BLASTP similarity searches. The origin of each ORF was assigned according to criteria used in ref. 14, with ORFs listed in the order of decreasing alien index (AI). Color coding is the same as in *A*. (*C*) Presence of A and B lineages in five bdelloid species. Full-length copies that could be included into Fig. 2 are denoted by open plus signs in boldface type, whereas others are represented by partial fragments and could not be incorporated into the phylogeny. n.d., not detected.

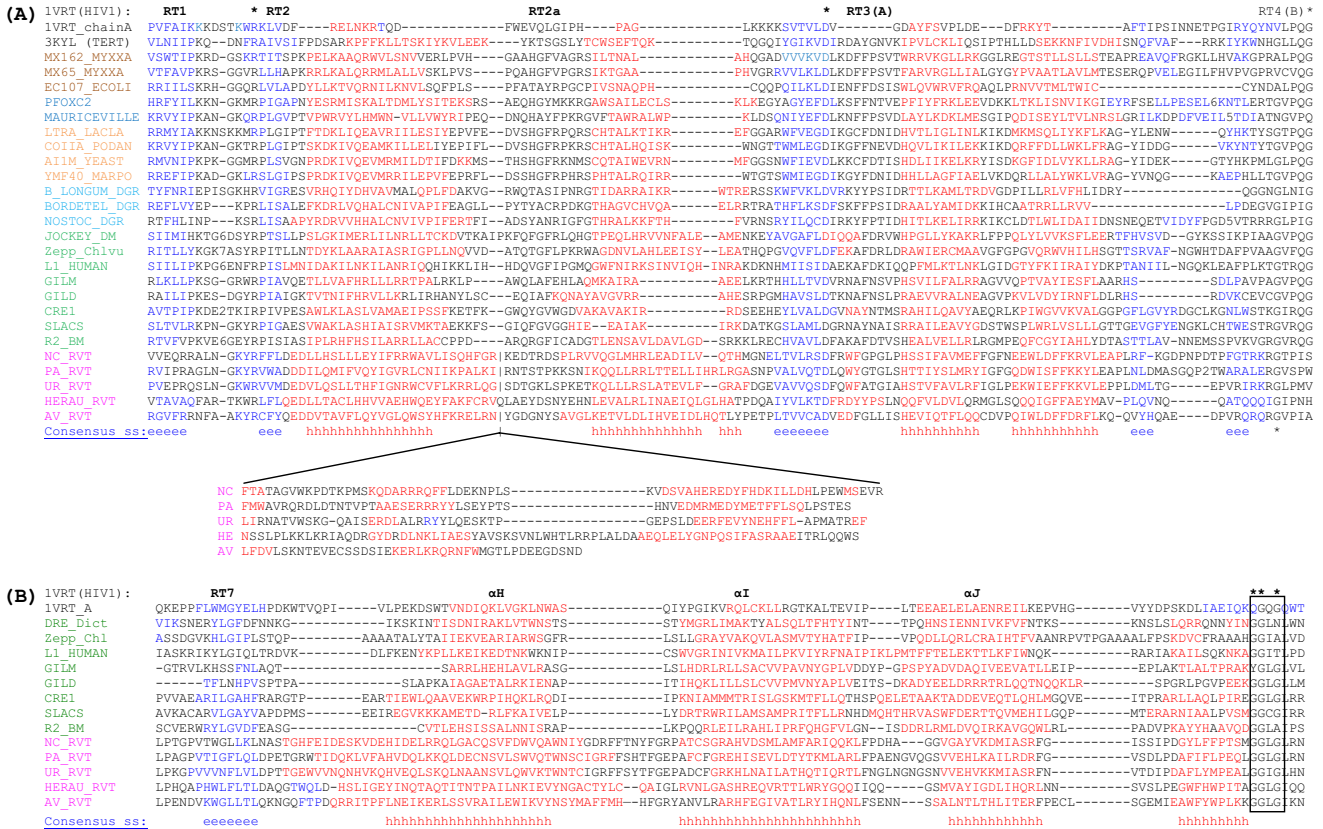
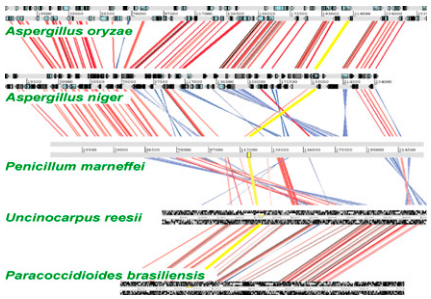


Fig. S2. (*A* and *B*) Partial structure-based alignments of the core RT domain including conserved motifs 1–4 and the *rvt* 2–3 loop region (*A*) and of the thumb subdomain (*B*). Shown are the RTs from retroviruses and TERT with known structures (from HIV-1 and *Tribolium castaneum*, in black); retrons (from *Myxococcus xanthus* and *E. coli*, in brown); retroplasmids (from *Fusarium oxysporum* and *N. crassa*, in blue); group II introns (from *Lactococcus lactis*, *Podospora anserina*, *Saccharomyces cerevisiae*, and *Marchantia polymorpha*, in orange); diversity-generating retroelements (DGR) (from *Bifidobacterium longum*, *Bordetella bronchiseptica*, and *Nostoc punctiforme*, in cyan); non-LTR retrotransposons (*Drosophila melanogaster* jockey, *Dictyostelium discoideum* DRE, *Chlorella vulgaris* Zepp, *Homo sapiens* L1, *Giardia intestinalis* GilM and GilD, *Crithidia fasciculata* CRE1, *Trypanosoma brucei* SLACS, and *Bombyx mori* R2, in green); and *rvt* from *N. crassa, P. anserina, Uncinocarpus reesii, Herpetosiphon aurantiacus*, and *Adineta vaga* (in purple). Red and blue amino acids in the alignment symbolize α-helices and β-strands, respectively, and the consensus secondary structure predictions are shown in the bottom line. Vertical lines mark the position of the *rvt* loop domain inserted below. Highly conserved residues are denoted by asterisks; the GGLG motif common to *rvt* and LINE elements is boxed. The top lines summarize the 3D structure of HIV-1 RT with the α-helices in the thumb domain designated as in Huang et al. (15) and TcTERT as revealed by crystallographic analysis (16). Secondary structures of the remaining sequences were predicted by PSIPRED, as implemented on the PROMALS3D server (4) and verified on the Jpred 3 server (7).

A.



B.

Pairs of species	d_S	d_N	d_N/d_S
Aspergillus fumigatus Ts *Neosartorya fischeri Ts*	0.1501	0.0282	0.1877
Aspergillus oryzae Pa *Aspergillus terreus Pa*	11.202	0.3335	0.0298
Aspergillus nidulans Mo *Aspergillus terreus Mo*	27.062	0.2648	0.0098
Botryotinia fuckeliana Mo *Sclerotinia sclerotiorum Mo*	1.2257	0.1815	0.1481
Botryotinia fuckeliana Nc *Sclerotinia sclerotiorum Nc*	0.8693	0.1189	0.1368
Gibberella moniliformis Mo *Fusarium oxysporum Mo*	0.4831	0.0395	0.0817
Gibberella zeae Mo *Fusarium oxysporum Mo*	2.2753	0.1751	0.0770
Nectria haematococca Mo *Fusarium oxysporum Mo*	10.372	0.2450	0.0236
Neurospora crassa *Chaetomium globosum Nc*	67.877	0.3590	0.0053
Fomitiporia mediterranea 1 *Fomitiporia mediterranea 2*	0.5043	0.0913	0.1811
Fomitiporia mediterranea 2 *Fomitiporia mediterranea 3*	1.8873	0.2378	0.1260
Phytophthora ramorum *Phytophthora sojae*	1.2757	0.1132	0.0887
Phytophthora capsici *Phytophthora sojae*	2.1919	0.1517	0.0692
Philodina roseola *Macrotrachela quadricornifera*	1.1110	0.1072	0.0965

Fig. S3. Synteny in *rvt* genomic environments and purifying selection in copies from syntenic regions. (*A*) Syntenic regions surrounding *rvt* from Ts lineage in five Eurotiomycetes. Shown are the results of four pairwise TBLASTX comparisons between *rvt*-containing contigs, visualized with the Artemis Comparison Tool (ACT). Red, TBLASTX hits in direct orientation; blue, those in inverted orientation; *rvt* genes are connected by yellow lines. (*B*) Evidence of purifying selection acting on *rvt* genes in related species. For *rvt* lineage designation, see text and Fig. 2. Synonymous (d_S) and nonsynonymous (d_N) substitution rates per site were calculated for 14 pairwise rvt combinations (*Materials and Methods*).

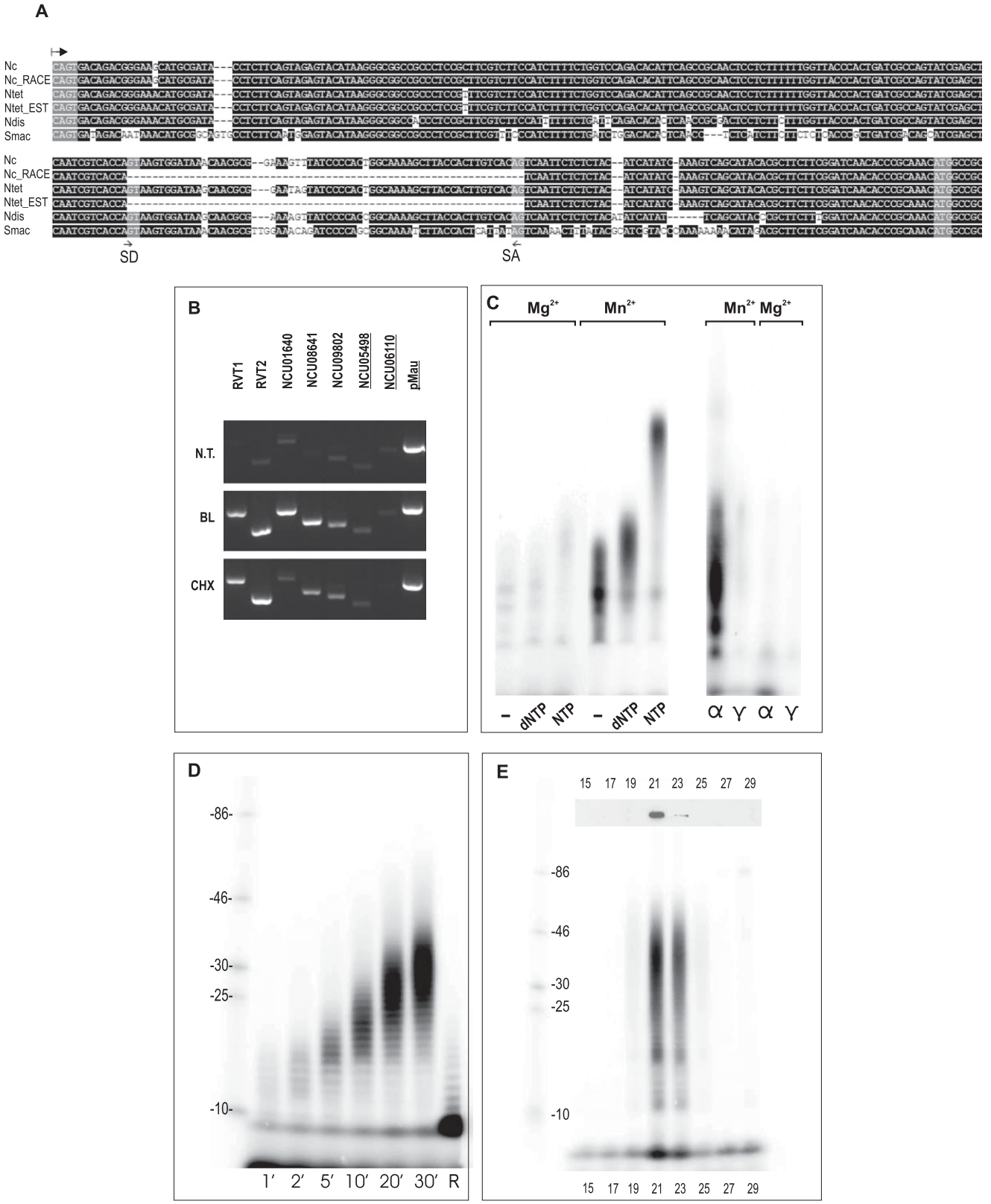


Fig. S4. Analysis of *N. crassa rvt* expression and activity. (*A*) Intron conservation in the 5′-UTR of *rvt* from *N. crassa*, *N. tetrasperma*, *N. discreta*, and *S. macrospora*. Genomic sequences are aligned with a sequenced 5′-RACE product from *N. crassa* and with an EST cluster consensus (Joint Genome Institute) from *N. tetrasperma*. Transcription start site, splice donor and acceptor sites, and the ATG codon are highlighted. (*B*) Semiquantitative RT-PCR analysis of mRNA levels after induction with blasticidin (BL) and cycloheximide (CHX). RVT1 and RVT2 are different *rvt* primer combinations (*Materials and Methods*). Control genes that do not respond to either treatment (NCU05498, NCU06110, and pMau) are underlined. NCU01640 is a C2H2 transcriptional regulator of the 26S proteasome (up-regulated by both antibiotics). NCU08641, AAA^+ ATPase (up-regulated by blasticidin); NCU09802, hypothetical 99-aa protein (up-regulated by

Legend continued on following page

both). These genes were identified in a preliminary microarray experiment. (*C*) NTP and dNTP addition by the His-tagged NcRVT in the presence of Mg^{2+} and Mn^{2+}. (*Left*) NcRVT was incubated with [α-^{32}P]dATP in the presence of 3 mM Mg^{2+} or 3 mM Mn^{2+} and then chased with a mix of four dNTPs or NTPs. (*Right*) Initial incorporation of [α-^{32}P]dATP (α) or [γ-^{32}P]dATP (γ) at the labeling step in the presence of 3 mM Mn^{2+} or 3mM Mg^{2+} as indicated. (*D*) Time course of NTP incorporation by the His-tagged NcRVT. Size markers (nt) are indicated. The right lane shows an ~10-nt fragment protected from RNase digestion, possibly by RVT. (*E*) Copurification of His-tagged NcRVT with terminal nucleotidyltransferase activity after sucrose gradient fractionation. Fractions were assayed for RVT presence by Western blotting (*Upper*) and for NTP incorporation as in *C*. Fractions are numbered from the bottom.

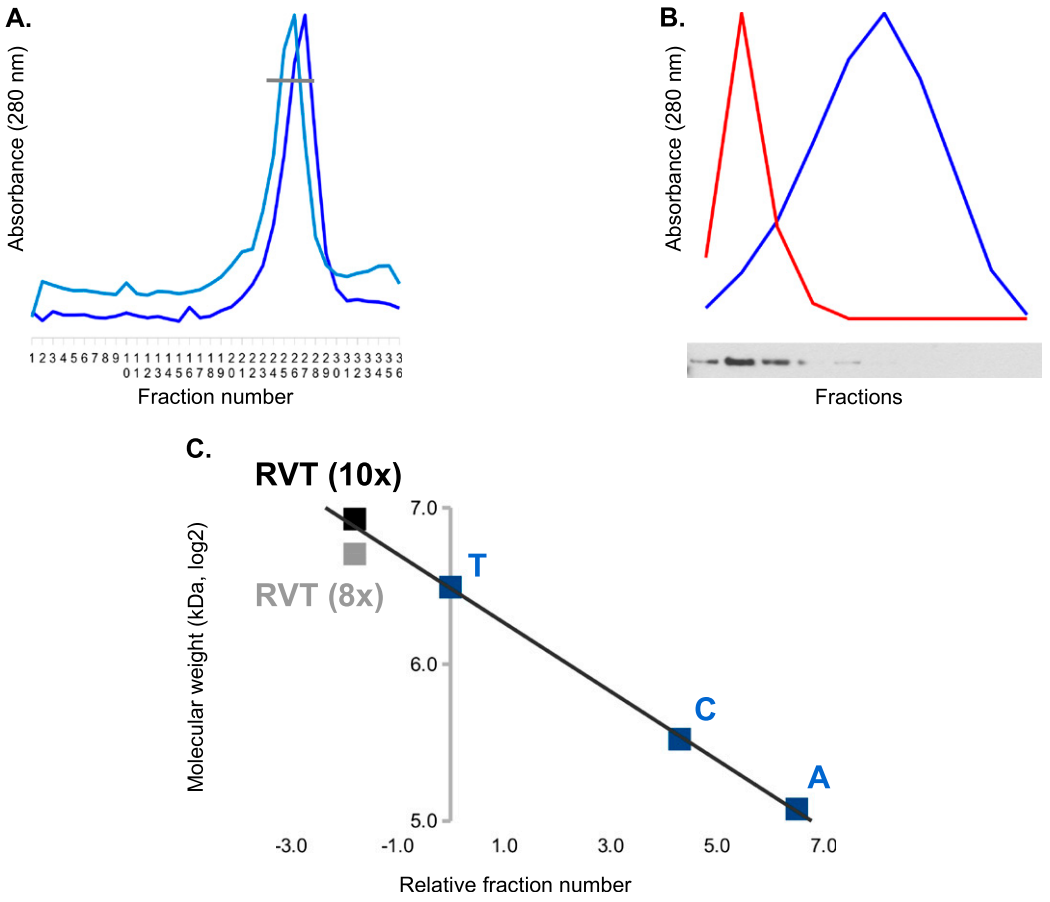


Fig. S5. Determination of the approximate molecular weight of the NcRVT complex. (*A*) Two representative sucrose gradients with thyroglobulin. Gray bar, sucrose fractions assayed for the presence of NcRVT by Western blot analysis. (*B*) Sucrose gradient with thyroglobulin and 6xHis-NcRVT (WT). Blue graph, 280-nm absorbances of sucrose fractions; red graph, normalized Western blot signal intensities, corresponding to the Western blot shown below. (*C*) Relative position of the NcRVT complex to the three molecular weight standards, T (thyroglobulin, 660 kDa), C (catalase, 250 kDa), and A (aldolase, 160 kDa).

Match to: **gi|38566868** Score: **293** Expect: **1.2e-23**
conserved hypothetical protein [Neurospora crassa]

Nominal mass (M_r): **102676**; Calculated pI value: **5.60**
NCBI BLAST search of gi|38566868 against nr
Taxonomy: Neurospora crassa
Links to other entries containing this sequence from NCBI Entrez:
gi|32418834 from Neurospora crassa
gi|28919200 from Neurospora crassa

Fixed modifications: Carbamidomethyl (C)
Variable modifications: N-Acetyl (Protein), Oxidation (M),Pyro-glu (N-term Q)
Cleavage by Trypsin: cuts C-term side of KR unless next residue is P
Number of mass values searched: **93**
Number of mass values matched: **64**
Sequence Coverage: **61%**

Matched peptides shown in **Bold Red**

```
    1 MAASSEVLSQ TLSSITSIKL DQLQKQKDAY ESAKDALLSA ADKEADVRKR
   51 AETLLDGREK LPSIRRADNP MLSADNMKRF VEQAAFDPSV SKDLLREYEE
  101 TVKKELDMTS NKYRFASMYG RMVREWTAAS GQDKKMSVTE GGDKADKDDF
  151 VPVGRKEMHE QRKTWEEFVF TPKETDKDAI KRYLEDVFAG SSKDCKRALA
  201 ELRKSFEELQ DNKHANWSHP FTVLQVKVCI QSILRSNNIT GEKRSTLRDF
  251 LNNSVVLQEI ADVLNMRMDA RASWTWEAPV VVEQRRALNG KYRFFLDEDL
  301 LHSLLLEYIF RRWAVLISQH FGRFTATAGV WKPDTKPMSK QDARRRQFFL
  351 DEKNPLSKVD SVAHEREDYF HDKILLDHLP EWMSEVRGGY DSSEADSKED
  401 TRDSPLRVVQ GLMHRLEADI LVQTHMGNEL TVLRSDFRWF GPGLPHSSIF
  451 AVMEFFGFNE EWLDFFKRVL EAPLRFKGDP NPDTPFGTRK RGTPISSTIS
  501 DVVGESLLFC LDFAVNQKAD GTLLYRLHDD MWLWGTTEKC SKAWKVVTEF
  551 SEVMGLSLNE EKTGSATIHP KNKKVGLEST KEEETAKISH NLPTGPVTWG
  601 LLKLNASTGH FEIDESKVDE HIDELRRQLG ACQSVFDWVQ AWNIYGDRFF
  651 TNYFGRPATC SGRAHVDSML AMFARIQQKL FPDHAGGVGA YVKDMIASRF
  701 GISSIPDGYL FFPTSMGGLG LRNPFVSLFL IRDDLEKTPE EMLADYEEEE
  751 ERAYRRAKER FETYEMERAV NKTAGGSTRG QDSFKDLEGE PFMSYEEFTR
  801 YRELTSPRLK AFYQNLLMEP KTKNVELKGD IKAALEDEDD WEDMSSYDKW
  851 VVQLFHREVV DMFGGLTVVD KAALPIGLIT MLRQSRFHWQ G
```

Fig. S6. Mass-spectrometry results for the purified *N. crassa* RVT protein.

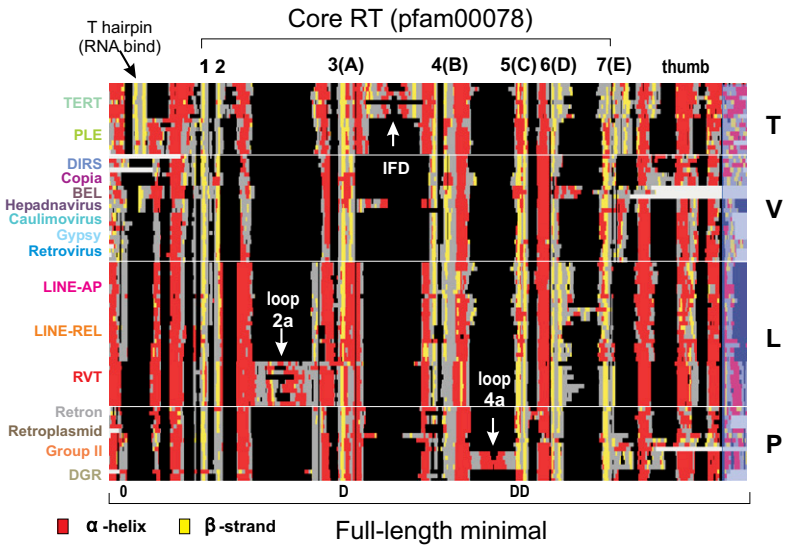


Fig. S7. An overview of structure-based alignment showing RT secondary structure elements in different colors (α-helices, red; β-sheets, yellow; random coils, gray). Visualization was assisted by the program STRAP (www. bioinformatics.org/strap). White arrows point to group-specific insertions between motifs, such as TERT IFD (insertion in the fingers domain), or *rvt* loop between motifs 2 and 3. The characteristic β-hairpin common to TERT and PLE (the so-called T domain) is indicated by a black arrow; another region of similarity immediately follows motif 7. The shortest RTs, such as retrons, retroplasmids, and DGRs, usually do not contain any additional N- or C-terminal extensions.

Table S1. Diversity of *rvt* sequences in public databases

Species	Taxonomy	Genomic copy No.	*rvt*-related EST	Total EST
Allomyces macrogynus	Blastocladiomycetes (Chytrids)	3	1	5,082
Acremonium alcalophilum	Ascomycetes; Sordariomycetes	1	n/a	0
Ajellomyces capsulatus	Ascomycetes; Eurotiomycetes	2	2	26,389
Ajellomyces dermatitidis	Ascomycetes; Eurotiomycetes	2	n/a	0
Alternaria brassicicola	Ascomycetes; Dothideomycetes	1	1	10,688
Aspergillus carbonarius	Ascomycetes; Eurotiomycetes	2	50+32	2,466,582
Aspergillus clavatus	Ascomycetes; Eurotiomycetes	2	n/a	0
Aspergillus flavus	Ascomycetes; Eurotiomycetes	3	2	20,371
Aspergillus fumigatus	Ascomycetes; Eurotiomycetes	2	0	180
Aspergillus nidulans	Ascomycetes; Eurotiomycetes	3	1	16,848
Aspergillus niger	Ascomycetes; Eurotiomycetes	2	1	46,938
Aspergillus oryzae	Ascomycetes; Eurotiomycetes	3	2+2	9,051
Aspergillus terreus	Ascomycetes; Eurotiomycetes	3	n/a	0
Botryotinia fuckeliana	Ascomycetes; Leotiomycetes	2	0	10,982
Chaetomium globosum	Ascomycetes; Sordariomycetes	2	1	1,557
Coccidioides immitis	Ascomycetes; Eurotiomycetes	2	n/a	0
Coccidioides posadasii	Ascomycetes; Eurotiomycetes	2	5+1	53,664
Cochliobolus heterostrophus	Ascomycetes; Dothideomycetes	1	2	88,751
Cryphonectria parasitica	Ascomycetes; Sordariomycetes	1	0	22,917
Fusarium oxysporum	Ascomycetes; Sordariomycetes	1	0	9,248
Geomyces destructans	Ascomycetes; Leotiomycetes	1	n/a	0
Gibberella moniliformis (F. verticillioides)	Ascomycetes; Sordariomycetes	2	2	87,086
Gibberella zeae (Fusarium graminearum)	Ascomycetes; Sordariomycetes	1	1	21,355
Glomerella graminicola	Ascomycetes; Sordariomycetes	1	0	2,380
Leptosphaeria maculans	Ascomycetes; Dothideomycetes	2	0	1,325
Magnaporthe grisea	Ascomycetes; Sordariomycetes	1	18	88,292
Metarhizium anisopliae	Ascomycetes; Sordariomycetes	1	n/a	n/a
Metarhizium acridum (fragment)	Ascomycetes; Sordariomycetes	1	n/a	n/a
Myceliophthora thermophila	Ascomycetes; Sordariomycetes	2	3	44,939
Nectria haematococca	Ascomycetes; Sordariomycetes	2	n/a	33,142
Neosartorya fischeri	Ascomycetes; Eurotiomycetes	1	n/a	0
Neurospora crassa	Ascomycetes; Sordariomycetes	1	23	277,147
Neurospora discreta	Ascomycetes; Sordariomycetes	1	7	48,084
Neurospora tetrasperma	Ascomycetes; Sordariomycetes	1	35	279,323
Paracoccidioides brasiliensis	Ascomycetes; Eurotiomycetes	2	n/a	0
Pyrenophora teres f. teres	Ascomycetes; Dothideomycetes	1	n/a	0
Pyrenophora tritici-repentis	Ascomycetes; Dothideomycetes	2	n/a	0
Penicillium marneffei	Ascomycetes; Eurotiomycetes	1	0	43
Penicillium chrysogenum	Ascomycetes; Eurotiomycetes	1	0	107
Phaeosphaeria nodorum	Ascomycetes; Dothideomycetes	4	0	15,973
Podospora anserina	Ascomycetes; Sordariomycetes	1	n/a	0
Sclerotinia sclerotiorum	Ascomycetes; Leotiomycetes	2	1	1,494
Sordaria macrospora	Ascomycetes; Sordariomycetes	1	n/a	0
Talaromyces stipitatus	Ascomycetes; Eurotiomycetes	1	n/a	0
Thielavia (Myceliophthora) terrestris	Ascomycetes; Sordariomycetes	2	6	27,991
Tuber melanosporum	Ascomycetes: Pezizomycetes	2	2	7,895
Uncinocarpus reesii	Ascomycetes; Eurotiomycetes	2	n/a	0
Verticillium albo-atrum	Ascomycetes; Sordariomycetes	2	5	20,813
Verticillium dahliae	Ascomycetes; Sordariomycetes	2	1	2,502
Vavraia culicis	Microsporidia; Pansporoblastina	1	n/a	0

Table S1. Cont.

Amanita bisporigera	Basidiomycetes; Agaricomycetes	2	n/a	0
Coprinopsis cinerea (fragment)	Basidiomycetes; Agaricomycetes	1	n/a	0
Fomitiporia mediterranea	Basidiomycetes; Agaricomycetes	3	12+148+45	1,287,882
Laccaria bicolor	Basidiomycetes; Agaricomycetes	1	4	34,335
Pleurotus ostreatus	Basidiomycetes; Agaricomycetes	1	4	29,116
Schizophyllum commune	Basidiomycetes; Agaricomycetes	1	0	31,336
Ustilago maydis	Basidiomycetes; Agaricomycetes	1	3	39,308
Phytophthora infestans	Stramenopiles; Oomycetes	1	1	94,091
Phytophthora sojae	Stramenopiles; Oomycetes	1	0	28,357
Phytophthora ramorum	Stramenopiles; Oomycetes	1	n/a	0
Phytophthora capsici	Stramenopiles; Oomycetes	1	1	56,457
Saprolegnia parasitica	Stramenopiles; Oomycetes	2	n/a	0
Pythium ultimum	Stramenopiles; Oomycetes	1	0	100,391
Physcomitrella patens	Viridiplantae; Streptophyta	1	32	326,059
Herpetosiphon aurantiacus	Bacteria; Chloroflexi	1	n/a	0
Environmental traces	Bacteria	2	n/a	0
Chaetomium cupreum (EST)	Ascomycetes; Sordariomycetes	n/a	1	4,285
Coniothyrium minitans (EST)	Ascomycetes; Dothideomycetes	n/a	2	602
Lentinula edodes (EST)	Basidiomycetes; Agaricomycetes	n/a	1	12,144
Phanerochaete carnosa (EST)	Basidiomycetes: Agaricomycetes	n/a	2	SRA
Onychiurus arcticus (EST)	Arthropoda;Hexapoda; Collembola	n/a	2	16,379

Plus sign between EST numbers indicates expression from different lineages. Species names are color coded according to taxonomy. Five EST matches have not yet been confirmed by genomic sequencing, although most ESTs have been upgraded to genomic sequences upon completion of genome projects.

Other Supporting Information Files

Dataset S1 (TXT)