Supporting Online Material

Materials and Methods

Sample collection

Samples of adult bees and combs were collected from migratory beekeeping operations experiencing CCD. Operations were chosen with different home bases and pollination routes. Diseased apiaries were selected based on evidence of recent collapse of some colonies within the apiary and a lack of dead bees in collapsed colonies. Up to three dead, collapsing or stronger colonies were selected for sample collection in each diseased apiary. In living colonies, samples of approximately 300 adult bees were collected in 70% ethanol; approximately 150 adult bees were frozen on dry ice and stored at -80°C until processed. All adult bee samples were collected from two brood nest combs shaken into a plastic tub. Similar samples were removed from control apiaries (that were in apparent good health and were < 3 km from CCD affected apiaries), apiaries far from any reported collapse in Hawaii, historical samples (2004) from Pennsylvanian colonies with known multi-year survivorship, and non-CCD colonies in Arizona (2006), Georgia (2007) and Louisiana (2004). Samples from imported Australian bees were collected directly from shipping packages. Samples collected in Florida included colonies that had migrated from Pennsylvania (CCD) and Georgia (non-CCD) in the previous three months. Samples collected from Californian apiaries represent colonies that were migratory within California and that had migrated from Minnesota, Montana and Washington in the last three months. All apiaries sampled in California contained collapsing colonies. Royal jelly samples (four separate lots) were purchased from a U.S. distributor that had imported royal jelly from northern Manchuria in China. Due to import restrictions, bees could not be obtained from China. Some

queen breeders in the U.S. and in other countries use purchased royal jelly to wet-graft and promote queen production.

Parasitic mite and Nosema analysis

Adult bee samples in alcohol were used to determine the prevalence of *Acarapis woodi*, *Varroa destructor* and *Nosema* spp., following standard methods developed by the Office International des Épizooties (*1*). *A. woodi* were detected in 24% of all samples and at low levels. Varroa mites were found in approximately 50% of samples but mite levels between CCD and healthy colonies did not differ. *Nosema* spp. were detected in approximately 50% of the samples and, again, the levels were similar between apparently healthy and CCD affected colonies (Table S2).

Sample preparation

Individual samples composed of 60 bees from the same operation and location (10 bees from six colonies) were ground in liquid nitrogen and immediately extracted using Tri-Reagent (Molecular Research Center Inc., Cincinnati, OH). Four royal jelly samples were extracted using Tri-Reagent. First-strand reverse transcription was initiated with a random octamer linked to a specific primer sequence (5'-GTT TCC CAG TAG GTC TCN NNN NNN N-3') (*2*). After digestion with RNase H, cDNA was amplified by using a 1:9 mixture of the above primer and a primer targeting the specific primer sequence (5'-CGC CGT TTC CCA GTA GGT CTC-3'). Initial PCR amplification cycles were performed at a low annealing temperature (25°C); subsequent cycles used a stringent annealing temperature (55°C) to favor priming through the specific sequence. Pooled samples were submitted for high-throughput pyrosequencing at 454 Life Sciences (Branford, CT).

Automated sequence analysis

Sequences ranged in size from 40 to 400 base pairs (bp) (150 bp average). Raw sequence reads were trimmed to remove sequences derived from the amplification primer and further filtered to eliminate highly repetitive sequences. After trimming and elimination of repeats, sequences were then clustered into non-redundant sequence sets using CD-HIT (*3*). Unique sequence reads were assembled into contiguous sequences (contigs) using the CAP3 algorithm (*4*). The quality scores for contigs and singlets were used to mask all positions with a phred score of less than 20. Contigs and unassembled singlet sequences were compared to the non-redundant sequence databases at NCBI (http://www.ncbi.nlm.nih.gov) using BLASTN (nucleotide homology) and BLASTX (protein homology) (*5*). The resulting BLAST alignments were analyzed and assigned to nodes in the NCBI taxonomy database using a custom application written in Perl/Bioperl. 16S sequences were identified by alignment to a limited set of 5185 type-species 16S rRNA sequences downloaded from the Ribosomal Database Project II (http://rdp.cme.msu.edu/) (*6*).

Confirmation by PCR and conventional sequencing

In addition to using published diagnostic primer sets (*7*), primers were designed for all candidate pathogens based on the recovered sequences and related sequences in the database. Primers were designed using SCPrimer (*8*) or Primer3 (*9*), depending on the number of sequences available for a given pathogen in the metagenomic sequence data and public databases. Aliquots of each sample were subjected to PCR amplification with HotStar PCR kit (Qiagen, Valencia, CA), 0.5 µmol/L of each primer, and the following cycling protocol: an annealing step with a temperature reduction in 1°C increments from 65°C to 51°C during the first 15 cycles and then continuing with a cycling profile of 94°C for 60 seconds (sec), 50°C for 60 sec, and 72°C for 60 sec in an MJ PTC200 thermal cycler (MJ Research, Waltham, MA, USA). Amplified products from each PCR reaction

were submitted for conventional sequencing, to confirm the presence of the pathogen and recover sequence for phylogenetic analysis. Additional PCR assays were established to amplify phylogenetically informative regions of selected pathogens (KBV, IAPV, kinetoplastids, entomophthorales) using the primers detailed in Table S3.

Bacterial 16S gene sequences were amplified from individual samples from Australia, Hawaii, and the continental United States and included both CCD and non-CCD colonies. Tri-Reagent extracted DNA was amplified using broad-range bacterial 16S primers Bact-16S-774A (5'-GTA GTC CAC GCT GTA AAC GAT G-3') (*10*) and Bact-1391R (5'-GAC GGG CGG TGT GTR CA-3') (*11*). PCR was performed using the HotStar PCR kit (Qiagen) in a GeneAmp PCR System 9700 thermocycler (Applied Biosystems, Foster City, CA). The following cycling parameters were used: 12 minutes (min) of initial denaturation at 95ºC followed by 35 cycles of 95°C for 30 sec, 56°C for 30 sec, and 72°C for 60 sec and a final extension at 72ºC for 8 min. Amplified products were cloned using the pGEM-T Easy Vector system (Promega, Madison, WI) and submitted for conventional sequencing using ABI Prism 3730xl DNA analyzers (Genewiz, Plainfield, NJ). Sequences were trimmed to remove vector sequences and classified by using Megablast to query GenBank. Sequences were categorized as corresponding to a previous bee commensal if the first blast hit was to a sequence previously obtained from a honey bee sample.

Quantitative PCR

RNA was reverse transcribed using random hexamer primers and SuperScript II Reverse Transcriptase (Invitrogen, Carlsbad, CA). SYBR Green Real-Time PCR was performed using primers described in Table S3; cDNA copy number was quantitated using a standard curve generated by known concentrations of the matching plasmid

standard. Template cDNA was mixed with the SYBR Green PCR Mix (Applied Biosystems, Foster City, CA), forward and reverse primers and incubated at 50°C, 2 min; 95°C, 10 min. SYBR Green signal was measured for duplicate samples and standards for 44 cycles: 95°C 15 sec; 60°C, 1 min. *A. mellifera* actin mRNA was used to assay RNA integrity.

Phylogenetic analysis

Agents identified in this study were compared to the nearest sequences recovered by BLAST. Bacterial sequences recovered by PCR were trimmed to remove vector sequences and classified using BLASTN. All homologous sequences were then aligned using either CLUSTALX (v1.83) (*12*) or MUSCLE (*13*). The highly divergent nature of the fungal 18S sequences necessitated exclusion of all regions where the alignment was ambiguous from the final data set used for phylogenetic analysis. Phylogenetic analysis was performed using the maximum likelihood (ML) method available in PAUP* (*14*) under TBR branch-swapping with the best-fit model of nucleotide substitution in each case determined using MODELTEST (*15*) (all parameter values available from the authors on request). To assess the reliability of specific groupings on the phylogenetic trees, a bootstrap resampling analysis was undertaken, employing 1000 replicate neighbor-joining trees under the ML substitution model.

Statistical Analysis

The prevalence of each pathogen in CCD and non-CCD samples, and potential associations between pathogens within samples, were examined by Chi-square analysis using StatView statistical software (v. 5.0.1, SAS Institute, Cary, NC). The positive predictive value (PPV), sensitivity and specificity associated with PCR assay results for

each pathogen were calculated for each of the four individual agents (IAPV, KBV, *N. apis*, *N. ceranae*) using data from 2x2 contingency tables. To determine whether samples from CCD as compared to non-CCD operations were more likely to contain multiple pathogens within single samples, we compared the number of pathogens found in each sample between the two groups using non-parametric statistics (Mann-Whitney U test) due to non-Gaussian distribution of data. Where the presence of multiple pathogens was found to be significantly correlated within either the total sample set or within CCD or non-CCD samples, multinomial logistic regression was used to examine the relationship of all four infectious agents, singly and in combination, to the presence of CCD. A Wald test of $p<0.05$ was defined as significant; main effects of independent variables on CCD outcomes were then examined for significance based on logistic likelihood ratio tests. Odds ratios and 95% confidence intervals were derived from logistic model coefficients tables. The significance of the whole model fit was assessed by Pearson and deviance residual goodness-of-fit methods and likelihood ratio tests, and the degree of variance explained by the logistic model was examined by comparing the correspondence of outcomes observed and those predicted by the model using logistic classification tables.

Models that included as the independent variable either the presence of *N. ceranae* as the only Nosema representative, or KBV as the only dicistrovirus in an individual sample, were associated with reduced risk for CCD (odds ratio << 1); these models also fit the data less well and had decreased explanatory power. A multinomial logistic regression model evaluating the following independent variables provided maximal fit to the data: 1) IAPV; 2) KBV; 3) *N. ceranae*; 4) *N. apis* (likelihood ratio for logistic whole model fit, $p<0.0001$; explanatory power of logistic classification, 88.2%).

Supporting Tables

Table S1. Results of SYBR Green quantitative PCR. Values are average RNA copies/bee, rounded to the nearest order of magnitude, for RNA extracted from pools of 60 bees (ND, not detected). [1] Each CCD operation is indicated by the geographic range. NC, *Nosema ceranae*; NA, *Nosema apis;* M, *Mucor* spp.; VD, *Varroa destructor* mite; CBPV, Chronic bee paralysis virus; SBV, Sacbrood virus; DWV, Deformed wing virus; ABPV, Acute bee paralysis virus; BQCV, Black queen cell virus; KBV, Kashmir bee virus; IAPV, Israeli acute paralysis virus of bees.

	Fungus / Microsporidian			**Mite**	**Virus**	**Iflavirus**		**Dicistrovirus**			
Sample	**NC**	**NA**	**M**	**VD**	**CBPV**	**SBV**	**DWV**	**ABPV**	**BQCV**	**KBV**	**IAPV**
CCD bees [1]											
PA/FL	$3x10^8$	$8x10^3$	ND	$1x10^4$	ND	$3x10^7$	$3x10^7$	$3x10^5$	$3x10^4$	$1x10^4$	$1x10^7$
CA	$8x10^7$	$1x10^4$	$3x10^5$	ND	ND	ND	$1x10^4$	ND	$3x10^3$	ND	$3x10^3$
MT/CA	$8x10^3$	$8x10^6$	ND	ND	ND	$3x10^3$	$1x10^4$	ND	$3x10^3$	$5x10^3$	$8x10^4$
WA/CA	$3x10^8$	$6x10^5$	$3x10^4$	$5x10^4$	$<3x10^3$	$3x10^5$	$9x10^7$	$1x10^4$	$3x10^6$	$4x10^8$	$5x10^5$
Imported bees											
AUS	$<3x10^3$	$8x10^8$	ND	ND	ND	$3x10^4$	ND	ND	$3x10^3$	ND	$8x10^3$
non-CCD bees											
PA (PSU)	ND	$<3x10^3$	$4x10^5$	$1x10^6$	ND	$3x10^7$	$4x10^8$	ND	$3x10^3$	ND	ND
HI	$2x10^8$	ND	ND	ND	ND	$3x10^6$	$3x10^3$	$3x10^3$	$1x10^4$	ND	ND
Imported Royal Jelly											
M1	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND
M6	ND	ND	ND	ND	ND	ND	ND	ND	$3x10^3$	ND	ND
DR	$8x10^3$	ND	ND	ND	ND	ND	ND	ND	ND	ND	$3x10^4$
DO	$1x10^6$	ND	$3x10^4$	ND	ND	ND	$8x10^3$	ND	ND	ND	$3x10^3$

Table S2. Levels of two parasitic mites and *Nosema* spores were measured in 41 CCD colonies and 21 healthy colonies using adult bee samples collected in January 2007 from twelve beekeeping operations, representing 8 states. From each colony, 16 adult bees were examined for *A. woodi*, ~300 bees washed for *V. destructor* and a 30 bee sample microscopically examined for *Nosema* levels. Values shown indicate a mean ± 1 SD and the range of values is shown in parentheses. *t*-test analysis revealed no significant differences between CCD and healthy colonies for any of these parasites or pathogens.

	non-CCD colonies	CCD colonies	*P* value (*t*-test)
Acarapis woodi # Bees with mites	0.71 ± 1.64 (0 - 6)	0.44 ± 0.97 (0 - 4)	0.48
Varroa. destructor Mites / 100 bees	0.02 ± 0.05 (0 - 0.23)	0.01 ± 0.04 (0 - 0.26)	0.63
Nosema sp. Spores / bee	$1.4x10^6 \pm 3.5x10^6$ (0 - $15.3x10^6$)	$1.7x10^6 \pm 2.8x10^6$ (0 - $11.0x10^6$)	0.75

Table S3. Primers used in this study. Primers were used in SYBR Green real time PCR (SYBR) or conventional PCR (PCR). * Indicates that the primer sequences are based on sequences recovered in this study.

Target	Primer name	Sequence	Use
ABPV	B3S2603_L15 *	CAGTGTAGCTAGTTTAAAAGCCAATG	SYBR, PCR
ABPV	B3S2603_R164 *	AACATGCAGATTGAGACAGTTGA	SYBR, PCR
A. mellifera	actin-L	AGGAATGGAAGCTTGCGGTA	SYBR, PCR
A. mellifera	actin-R	AATTTTCATGGTGGATGGTGC	SYBR, PCR
BQCV	B2C151_L258 *	TTTGCTATGCGTGTGTAGGC	SYBR, PCR
BQCV	B2C151_R504 *	TGCGTAAATTGGTTCCATCA	SYBR, PCR
CBPV	L67	TTYGCTCGYTTCGAYCARACTCTGATG	SYBR
CBPV	R467	GCACGAAAGCCGAGRCARGADAG	SYBR
DWV	B1C118_L2 *	CAGTAGCTTGGGCGATTGTT	SYBR, PCR
DWV	B1C118_R171 *	AGCTTCTGGAACGGCAGATA	SYBR, PCR
Entomophthorales	F53	GCCAGTAGTCATATGCTTGTCTCAA	SYBR
Entomophthorales	R1080	ATGGTTGAGACTACGACGGTATCTA	SYBR
IAPV	B4S0427_L17 *	CGAACTTGGTGACTTGAAGG	SYBR, PCR
IAPV	B4S0427_R130 *	GCATCAGTCGTCTTCCAGGT	SYBR, PCR
IAPV	IAPV-F-RdRp	CGTCGACCCATTGAAAAAGT	PCR
IAPV	IAPV-R-RdRp	GGTTGGCTGTGTGTCATCAT	PCR
KBV	C_L34	CAAACTGCTGAATCAATGTCAAAAT	SYBR
KBV	C_R156	ACATGCCTCTACTTTGTCACATTCA	SYBR
KBV	KBV-F-RdRp	GATGAACGTCGACCTATTGA	PCR
KBV	KBV-R-RdRp	TGTGGGTTGGCTATGAGTCA	PCR
Trypanosomatidae	B_R106	GTTGACGGAATCAACCAAACAAAT	PCR
Trypanosomatidae	A_L40	GCGTCAGAGGTGAAATTCTTAGACC	PCR
N. apis	L203	CAGTTATGGGAAGTAACATAGTTG	SYBR
N. apis	R253	CGATTTGCCCTCCAATTAATCTG	SYBR
N. ceranae	L203	CAGTTATGGGAAGTAATATTATATTG	SYBR
N. ceranae	R253	TTGATTTGCCCTCCAATTAATCAC	SYBR
SBV	D_L29	GGTGTCTAACTTTTATGGACCACCA	SYBR
SBV	D_R132	CCTTTTCTATGCTATCATCCATCTGA	SYBR
Varroa mite	F24	TAACTGTTTTTGCCGACTTCATTGC	SYBR
Varroa mite	R396	CACTTAGTCGTAAGACATACGTAAGCA	SYBR
IAPV	IAPV_IGR_F	CGATGAACAACGGAAGGTTT	PCR
IAPV	IAPV_IGR_R	ATCGGCTAAGGGGTTTGTTT	PCR
KBV	KBV_IGR_F	GCCTAATTGGTGTCGAGGAG	PCR
KBV	KBV_IGR_R	GCTTTTCCACCAGCTTTCAA	PCR

Supporting Figures

Figure S1. (A) Bee sampling strategy. (I) RNA was extracted from ten bees collected from six individual colonies at each of seven operations to create a pool representing 60 bees per operation. (II) RNA from all operations was pooled for high throughput sequencing (HTS). (III) The presence of candidate pathogens found by HTS was investigated in the original 60 bee pools using various assays (ribosomal RNA consensus PCR, SYBR green real-time PCR, sequencing). (IV) Pools of 4 to 15 bees were analyzed for the presence of candidate pathogens using SYBR green real-time PCR or conventional PCR.

(A)

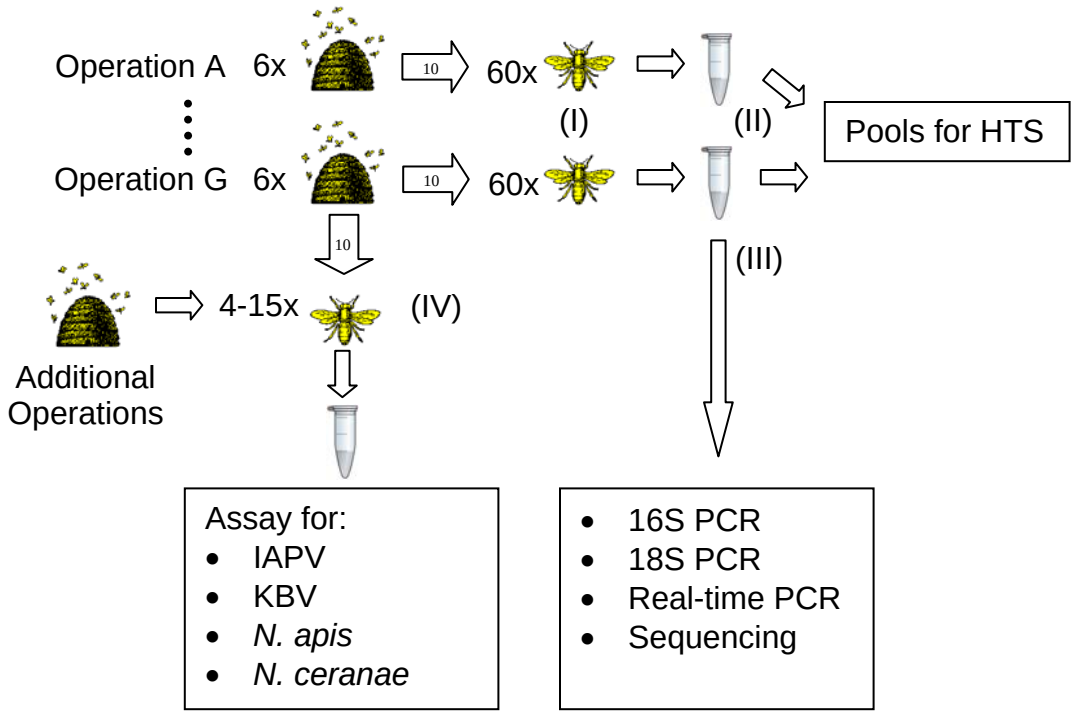


Figure S2. (B) High-throughput sequencing data were processed into contiguous sequences (steps 1-3). Short sequences that were 100% identical to a longer sequence were removed to form a non-redundant sequence set (red circles). Sequences were assembled by aligning overlapping sequences. BLAST was used to identify the closest relative in the NCBI sequence database (step 4).

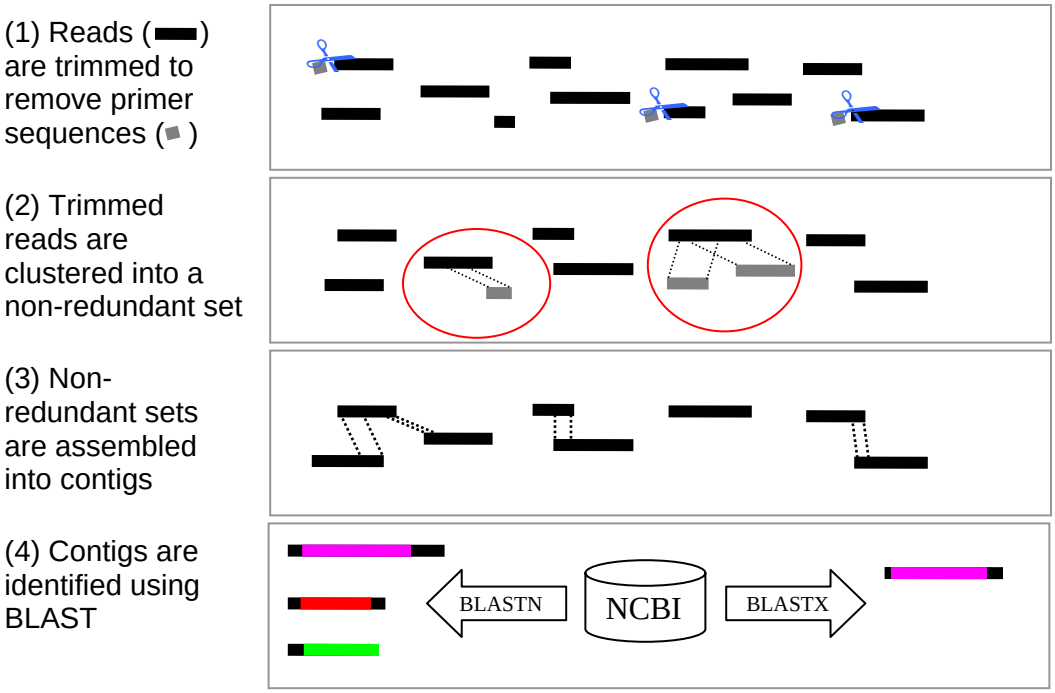
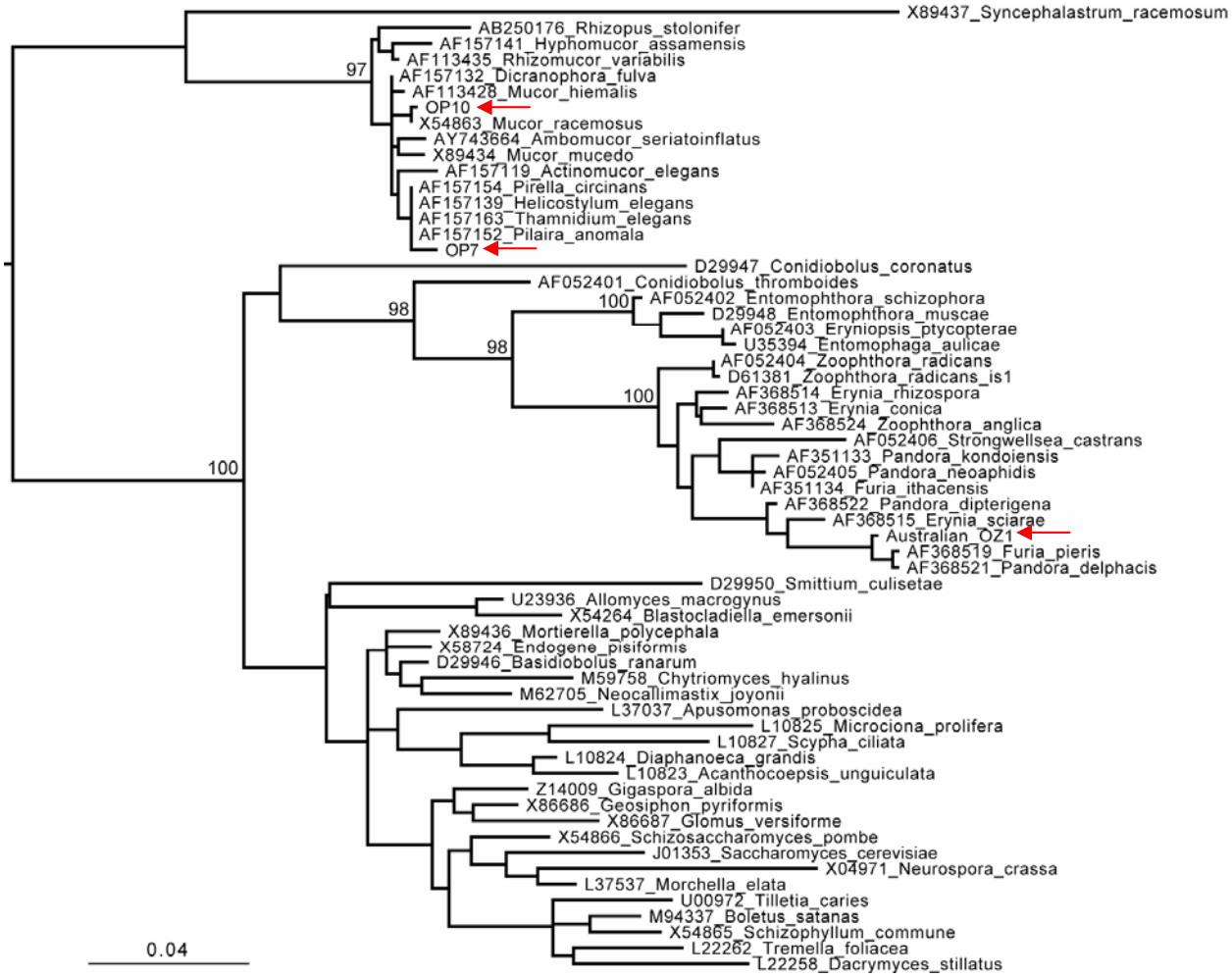


Figure S3. Maximum likelihood phylogenetic tree of 18S rRNA (715 nt) relating a fragment recovered from a CA sample (OP7) to other members of the *Trypanosomatidae*. All horizontal branch lengths are scaled according to the number of nucleotide substitutions per site. The tree is mid-point rooted for purposes of clarity only. Bootstrap values are shown for selected nodes.

Figure S4. Maximum likelihood phylogenetic tree of 18S rRNA fragments (722 nt) relating sequences recovered from bees in this study: WA/CA (OP10), CA (OP7), Australian (OZ1), to members of the Mucorales and Entomophthorales. All horizontal branch lengths are scaled according to the number of nucleotide substitutions per site. The tree is mid-point rooted for purposes of clarity only. Bootstrap values are shown for selected nodes.



References

1. *Manual of Standards for Diagnostic Tests and Vaccines*. (Office International des Épizooties, Paris, 2000).
2. D. Wang *et al.*, *Proc Natl Acad Sci U S A* **99**, 15687 (Nov 26, 2002).
3. W. Li, A. Godzik, *Bioinformatics* **22**, 1658 (Jul 1, 2006).
4. X. Huang, A. Madan, *Genome Res* **9**, 868 (Sep, 1999).
5. S. F. Altschul *et al.*, *Nucleic Acids Res* **25**, 3389 (Sep 1, 1997).
6. B. L. Maidak *et al.*, *Nucleic Acids Res* **29**, 173 (Jan 1, 2001).
7. M. Shen, X. Yang, D. Cox-Foster, L. Cui, *Virology* **342**, 141 (Nov 10, 2005).
8. O. J. Jabado *et al.*, *Nucleic Acids Res* **34**, 6605 (2006).
9. S. Rozen, H. Skaletsky, *Methods Mol Biol* **132**, 365 (2000).
10. P. H. Summanen *et al.*, *J Clin Microbiol* **43**, 4455 (Sep, 2005).
11. P. B. Eckburg *et al.*, *Science* **308**, 1635 (Jun 10, 2005).
12. R. Chenna *et al.*, *Nucleic Acids Res* **31**, 3497 (Jul 1, 2003).
13. R. C. Edgar, *Nucleic Acids Res* **32**, 1792 (2004).
14. D. L. Swofford. (Sinauer Associates Inc. Sunderland, MA, 2003).
15. D. Posada, K. A. Crandall, *Bioinformatics* **14**, 817 (1998).