Sciencexpress Report

A Metagenomic Survey of Microbes in Honey Bee Colony Collapse Disorder

Diana L. Cox-Foster,[1] Sean Conlan,[2] Edward C. Holmes,[3,4] Gustavo Palacios,[2] Jay D. Evans,[5] Nancy A. Moran,[6] Phenix-Lan Quan,[2] Thomas Briese,[2] Mady Hornig,[2] David M. Geiser,[7] Vince Martinson,[8] Dennis vanEngelsdorp,[1,9] Abby L. Kalkstein,[1] Andrew Drysdale,[2] Jeffrey Hui,[2] Junhui Zhai,[2] Liwang Cui,[1] Stephen K. Hutchison,[10] Jan Fredrik Simons,[10] Michael Egholm,[10] Jeffery S. Pettis,[5] W. Ian Lipkin[2*]

[1]Department of Entomology, The Pennsylvania State University, 501 ASI building, University Park, PA 16802, USA. [2]Center for Infection and Immunity, Mailman School of Public Health, Columbia University, New York, NY 10032, USA. [3]Center for Infectious Disease Dynamics, Department of Biology, The Pennsylvania State University, Mueller Laboratory, University Park, PA 16802, USA. [4]Fogarty International Center, National Institutes of Health, Bethesda, MD 20892, USA. [5]Bee Research Laboratory, USDA-ARS, Beltsville, MD 20705, USA. [6]Department of Ecology and Evolutionary Biology, University of Arizona, Tucson, AZ 85721, USA. [7]Department of Plant Pathology, The Pennsylvania State University, University Park, PA 16802, USA. [8]The Center for Insect Science, University of Arizona, Tucson, AZ 85721, USA. [9]The Pennsylvania Department of Agriculture, Bureau of Plant Industry - Apiculture, Harrisburg, PA 17110, USA. [10]454 Life Sciences, Branford, CT 06405, USA.

*To whom correspondence should be addressed. E-mail: wil2001@columbia.edu

In colony collapse disorder (CCD), honey bee colonies inexplicably lose their workers. CCD has resulted in a loss of 50 to 90% of colonies in beekeeping operations across the United States. The observation that irradiated combs from affected colonies can be repopulated with naive bees suggests that infection may contribute to CCD. We used an unbiased metagenomic approach to survey microflora in CCD hives, normal hives, and imported royal jelly. Candidate pathogens were screened for significance of association with CCD by examination of samples collected from several sites over a period of 3 years. One organism, Israeli acute paralysis virus of bees (IAPV), was strongly correlated with CCD.

Methods for cloning nucleic acids of microbial pathogens directly from clinical and environmental specimens afford unprecedented opportunities for pathogen discovery and surveillance. Subtractive cloning, polymerase chain reaction (PCR) and DNA microarrays have implicated novel pathogens as the etiological agents of several acute and chronic diseases. Here, we describe the application of unbiased high throughput pyrosequencing technology (*1*) in the characterization of the microflora associated with *Apis mellifera* in a search for the cause of colony collapse disorder (CCD).

CCD is characterized by the rapid loss from a colony of its adult bee population (*2–4*). No dead adult bees are found inside or in close proximity to the colony. At the end stages of collapse, a queen is attended only by a few newly emerged adult bees. Collapsed colonies often have considerable capped brood and food reserves. The phenomenon of CCD was first reported in 2006; however, beekeepers noted unique colony declines consistent with CCD as early as 2004. An estimated 23% of beekeeping operations in the USA suffered from CCD over the winter of 2006-2007. These beekeepers lost an average of 45% of their operations (*5*). Since the introduction of the varroa mite in the late 1980s, colonies have experienced increased mortality; however, in contrast to CCD, these colony deaths are marked by dead bees in the hive, an incremental decline in worker population, and robbing and pest invasion. One hypothesis is that CCD is due to the introduction of a novel infectious agent. This is supported by preliminary evidence that CCD is transmissible through reuse of equipment from CCD colonies and that such transmission can be broken by irradiation of the equipment before use (*6*).

Bees were analyzed in a metagenomic survey of four, widely separated operations across the USA affected by CCD. All operations were migratory, with wintering yards in either Florida or California in February 2007 [(*7*), fig. S1]. Two non-CCD samples were collected from Hawaii and Pennsylvania. An additional sample of apparently healthy bees imported from Australia, and four samples of imported royal jelly from China, were also tested as potential sources of pathogens. Total RNA was extracted to capture RNA viruses as well as other pathogens. The RNA was pooled as presumed CCD-positive, presumed CCD-negative, and royal jelly, for pyrosequencing. The raw sequencing reads were trimmed and assembled into contiguous sequences (contigs) (fig. S2). Analysis using BLASTN and BLASTX (*8*) revealed

the presence of bacteria, fungi, parasites, metazoa and viruses (Table 1).

Sequences homologous to bacterial 16S ribosomal RNA (16S rRNA) were assembled into 48 contigs. The majority (87%) of contigs aligned best to previously identified commensals of *A. mellifera* (Table 1). To obtain precision in typing the bacterial associates in the different samples, we obtained ~500 nt fragments of the 16S rRNA gene, using conserved, near-universal PCR primers. A total of 536 clones from the entire range of samples were characterized by conventional sequencing. When these sequences were used as queries in BLASTN searches of GenBank, over 96% gave closest hits to the eight clusters isolated from *A. mellifera* in previous studies (*9–11*). These clusters included an abundant member of the *Gammaproteobacteria* and several less frequent but widespread organisms from the *Betaproteobacteria*, *Alphaproteobacteria*, *Firmicutes* and *Actinobacteria* groups (Fig. 1). Only 20 bacterial 16S rRNA gene sequences were unidentified and they were not concentrated in CCD samples. Sequences of *Paenibacillus larvae* and *Streptococcus pluton*, the causative agents of American and European foulbrood, diseases of bee larvae, were not detected.

Bacterial analysis indicated community composition (Table 1) similar to that in samples collected in Africa, Switzerland and Germany (*9–11*), suggesting that *A. mellifera* has similar bacterial flora worldwide. Screens of the *A. mellifera* genome sequence database revealed sequences corresponding to several of these bacterial groups, indicating that they are likely part of the normal flora. The *A. mellifera* database includes sequences for 926 genes with significant reciprocal best BLAST hits to genes from completely sequenced bacterial genomes. The most abundantly represented groups corresponded to taxa in our samples, including 299 (32%) with best hits to *Lactobacillus* species and 249 (27%) with best hits to *Neisseria* species (members of the Betaproteobacteria and closest relatives of *Simonsiella* for which screening was possible based on the available full genome sequences). The gut lumen contains the majority of microorganisms in most insects. Because a similar profile of bacterial types was found in dissected intestines from *A. mellifera* (*9*, *11*), the bacterial species described here are likely to represent a characteristic gut-inhabiting community. Although we cannot exclude that a strain of a normally commensal bacterium has become pathogenic while retaining a near-identical 16S rRNA sequence, we observed no clear shift in abundance that would suggest this occurred in CCD. A trend towards increased abundance of one of the Gammaproteobacterial taxa in the CCD bees (Fig. 1) may reflect physiological changes accompanying CCD and impacting the commensal community. More sampling of individual colonies and age classes would be needed to determine the significance of this observation.

Metagenomic sequence data for a ~700 nt fragment of the 18S rRNA gene identified a trypanosomatid sequence. This sequence was identical across CCD and non-CCD samples. Phylogenetic analysis indicates that this parasite falls in the *Leptomonas-Leishmania-Crithidia* lineage, but a precise taxonomic assignment is not possible due to the paucity of rRNA data in this region (fig. S3) (*12–14*).

Eighty-one distinct fungal 18S rRNA sequences were recovered from the pooled samples, primarily from four distinct lineages: Saccharomycotina, which includes a variety of presumed commensal yeasts, Microsporidia, including the important bee pathogens *Nosema apis* and *N. ceranae* (*15*), Entomophtorales/Entomphthoromycotina, a diverse group of insect pathogens, and Mucorales/Mucoromycotina, which includes *Mucor hiemalis*, a species known to kill honeybees under certain conditions (*16*). We used 18S rRNA PCR primers designed to capture Entomophthorales and Mucorales (fig. S4). The occurrence of these fungi in our samples did not correlate with CCD (table S1). Although *N. ceranae* was detected by PCR in all operations affected by CCD, as well as the Australian sample and in two of the royal jelly samples, it was also detected in non-CCD samples. A specific PCR assay for *N. apis* showed that it was present in all 4 CCD operations as well as one non-CCD operation. *N. apis* was also found in the sample from Australia, but not in the royal jelly samples (table S1).

BLASTN analysis of the high-throughput sequencing data identified seven positive-sense single-stranded RNA viruses previously associated with disease in honey bees, including members of the family *Dicistroviridae* and the genus *Iflavirus*. The presence of the unclassified Chronic bee paralysis virus (CBPV) in only one of four CCD operations (table S1) suggests that it is not a primary contributor to this syndrome. Recovered sequences were used to establish specific quantitative PCR assays for the remaining six insect viruses. Two iflaviruses, Sacbrood virus (SBV) and Deformed wing virus (DWV), as well as two dicistroviruses, Black queen cell virus (BQCV) and Acute bee paralysis virus (ABPV), were found in both CCD and non-CCD operations.

Two other dicistroviruses, Kashmir bee virus (KBV) and Israeli acute paralysis virus of bees (IAPV), an unclassified virus that may reflect a distinct lineage of KBV or a new species, were found only in CCD operations. IAPV sequence analysis in the intergenic region (Fig. 2) indicated close phylogenetic relationship to both KBV and ABPV. IAPV was found in all four affected operations sampled, in two of four royal jelly samples and in the Australian sample (table S1). KBV was present in three of four CCD operations, but not in the royal jelly. KBV was not found in the sample of Australian bees shown in table S1; however, KBV was

subsequently found in five of eleven individual bees from the same Australian source.

Both *N. ceranae* and an unspecified iflavirus were proposed to be associated with CCD in an earlier report (*17*). We found *Nosema* spp. by PCR and spore count in both CCD and non-CCD operations (table S1, table S2), but no novel iflavirus. The overall prevalence of any Nosema species was 94.1% (100%, CCD; 85.7%, non-CCD). In contrast, the dicistroviruses KBV and IAPV correlated with CCD in our metagenomic survey. The prevalence of KBV, IAPV, *N. ceranae* and *N. apis* was tested in 51 pools of bees (4 to 15 bees per pool) collected from 30 CCD colonies and 21 non-CCD colonies between 2004 and 2007 in Arizona, California, Florida, Georgia, Louisiana, and Pennsylvania. Individual CCD samples were more likely than samples from non-CCD operations to contain more than one of these four pathogens. The mean number of pathogen types (± 1 standard deviation) found in individual samples from each site was 3.7 ± 0.5 for CCD samples versus 2.1 ± 0.9 for non-CCD samples ($p<0.0001$). The patterns of co-infection were complex and unevenly distributed throughout the sample set. All samples positive for IAPV contained KBV. Although KBV was prevalent in both CCD and non-CCD samples (90.2% of all samples), IAPV was, with a single exception, confined to CCD samples, yielding a positive predictive value of 96.1% and a specificity of 95.2% (Table 2). Multinomial logistic regression was pursued to determine the contributions of the four pathogens, singly and in combination, to CCD outcomes. Models with the best explanatory power included IAPV as one of the independent variables. IAPV was found to increase the risk of CCD (odds ratio=65; $p<0.0001$) with a trend for increased CCD risk in samples positive for *N. apis* (odds ratio=9; $p=0.053$). Neither KBV nor *N. ceranae* contributed significantly to the risk for CCD nor did they alter the influence of IAPV on CCD.

IAPV was first described in 2004 in Israel (*18*) where infected bees presented with shivering wings, progressed to paralysis and then died just outside the hive. All CCD operations sampled used imported bees from Australia or were intermingled with operations that had done so. Importation to the United States of bees from Australia began in 2004, coinciding with early reports of unusual colony declines. Although the shivering phenotype is not reported in imported Australian bees or in CCD, differences in IAPV pathogenicity may reflect strain variation, co-infection, or the presence of other stressors such as pesticides or poor nutrition. The varroa mite, for example, absent in Australia, immunosuppresses bees, making them more susceptible to infection by other organisms, including viruses (*19*, *20*). Other stressors may include chemical pesticides used on plants pollinated by bees and in hives to control pests. Crop pesticide use is similar in both the United States and Australia. Miticides are widely used in the United States but not in Australia, and can have adverse effects on colony health (*21*); however, miticide use did not differ between CCD and non-CCD operations as determined by detailed case histories (*22*).

We have used CCD as a model to establish a strategy for investigating epidemics of unexplained infectious disease. Metagenomic sequencing enabled rapid assembly of a comprehensive inventory of microflora in CCD and non-CCD populations and provided the foundation needed to address the significance and provenance of candidate pathogens. We have not proven a causal relationship between any infectious agent and CCD; nonetheless, the prevalence of IAPV sequences in CCD operations, as well as the temporal and geographic overlap of CCD and importation of IAPV-infected bees, indicate that IAPV is a significant marker for CCD.

References and Notes

1. M. Margulies *et al.*, *Nature* **437**, 376 (2005).
2. B. P. Oldroyd, *PLOS Biology* **5**, 1195 (2007).
3. A. Barrionuevo, “Bees vanish; Scientists race for reasons,” *New York Times*, April 24, 2007, p. 1.
4. E. Stokstad, *Science* **316**, 970 (2007).
5. D. vanEngelsdorp, R. Underwood, D. Caron, J. Hayes Jr., *Am. Bee J.* **147**, 599 (2007).
6. J. Pettis, D. VanEngelsdorp, D. Cox-Foster, *Am. Bee J.* **147**, 595 (2007).
7. Materials and methods are available as supporting material on *Science* Online.
8. S. F. Altschul *et al.*, *Nucleic Acids Res.* **25**, 3389 (1997).
9. D. Babendreier, D. Joller, J. Romeis, F. Bigler, F. Widmer, *FEMS Microbiol. Ecol.* **59**, 600 (2007).
10. A. Jeyaprakash, M. A. Hoy, M. H. Allsopp, *J. Invertebr. Pathol.* **84**, 96 (2003).
11. K. I. Mohr, C. C. Tebbe, *Environ. Microbiol.* **8**, 258 (2006).
12. S. Podlipaev, *Int. J. Parasitol.* **31**, 648 (2001).
13. S. A. Podlipaev *et al.*, *J. Eukaryot. Microbiol.* **51**, 283 (2004).
14. A. L. Hughes, H. Piontkivska, *Mol. Biol. Evol.* **20**, 644 (2003).
15. M. Higes, R. Martin, A. Meana, *J. Invertebr. Pathol.* **92**, 93 (2006).
16. C. E. Burnside, *Am. Bee J.* **25**, 75 (1935).
17. W. Ravven, “UCSF sleuths identify suspects in mystery of vanishing honeybees,” *UCSF Today*, April 25, 2007.
18. E. Maori, E. Tanne, I. Sela, *Virology* **362**, 342 (2007).
19. X. Yang, D. L. Cox-Foster, *Proc. Natl. Acad. Sci. U.S.A.* **102**, 7470 (2005).
20. P. G. Gregory, J. D. Evans, T. Rinderer, L. de Guzman, *J. Insect. Sci.* **5**, 7 (2005).

21. J. Pettis, A. Collins, R. Wilbanks, M. F. Feldlaufer, *Apidologie* **35**, 605 (2004).
22. D. vanEngelsdorp, D. Cox-Foster, M. Frazier, N. Ostiguy, J. Hayes Jr., "Fall-Dwindle Disease: A preliminary report" (CCD Working Group, 2006); available from http://maarec.cas.psu.edu/ColonyCollapseDisorderInfo.html.
23. We thank M. Andree, A. V. Bussetti, J. Chen, D. Grove, M. Hamilton, V. Levi, H. Lin, D. Lopez, S. McDonald, N. Rice, K. Roccasecca, B. Smith, O. Thompson, A. Ulsamer, V. Williams, and beekeepers who supplied samples used in these analyses. The work reported here was supported by National Institutes of Health awards U01AI070411, U54AI57158 (Northeast Biodefense Center-Lipkin), the National Honey Board, and the Pennsylvania Department of Agriculture.

Supporting Online Material
www.sciencemag.org/cgi/content/full/1146498/DC1
Materials and Methods
Figs. S1 to S4
Tables S1 to S3
References

14 June 2007; accepted 30 August 2007
Published online 6 September 2007;
10.1126/science.1146498
Include this information when citing this paper

Sciencexpress

Table 1. Closest sequenced relatives identified through BLAST analysis of the high-throughput sequence data. *Indicates viruses not yet classified by the International Committee on the Taxonomy of Viruses but that exhibit the key features of the indicated taxon. [1]Found in Jeyaprakash *et al*. (2003). [2]Found in Babendreier *et al*. (2007).

Kingdom	Taxon (rank)	Organism
Bacteria	*Firmicutes* (phylum)	*Lactobacillus* sp.[1,2] Uncultured Firmicutes[2]
Bacteria	*Actinobacteria* (class)	*Bifidobacterium* sp.[1]
Bacteria	*Alphaproteobacteria* (class)	*Bartonella* sp.[1,2] *Gluconacetobacter* sp.[1,2]
Bacteria	*Betaproteobacteria* (class)	*Simonsiella* sp.[1,2]
Bacteria	*Gammaproteobacteria* (class)	Two uncultured species[1,2]
Fungus	Entomophthorales (order)	*Pandora delphacis*
Fungus	Mucorales (order)	*Mucor* spp.
Fungus / Microsporidian	*Nosematidae* (family)	*Nosema ceranae*
Fungus / Microsporidian	*Nosematidae* (family)	*Nosema apis*
Eukaryota	*Trypanosomatidae* (family)	*Leishmania*/*Leptomonas* sp.
Metazoan	*Varroidae* (family)	*Varroa destructor*
Virus	(unclassified)	Chronic bee paralysis virus*
Virus	*Iflavirus* (genus)	*Sacbrood virus*
Virus	*Iflavirus* (genus)	Deformed wing virus*
Virus	*Dicistroviridae* (family)	*Black queen cell virus*
Virus	*Dicistroviridae* (family)	Kashmir bee virus*
Virus	*Dicistroviridae* (family)	*Acute bee paralysis virus*
Virus	*Dicistroviridae* (family)	Israeli acute paralysis virus of bees*

Table 2. Analysis of pools of bees tested for candidate pathogens. The positive predictive value represents the probability that a positive result for a given agent is associated with CCD. The sensitivity is the probability that test results will be positive in all CCD cases. Specificity is defined as the probability that all non-CCD samples will be associated with negative test results. IAPV, Israeli acute paralysis virus of bees; KBV, Kashmir bee virus.

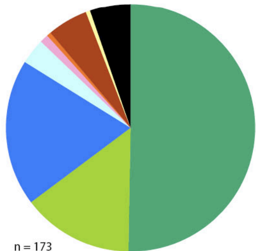
Agent	Number of positive samples *n* (% positive of samples tested)			Positive Predictive Value (%)	Sensitivity (%)	Specificity (%)
	CCD (n = 30)	non-CCD (n = 21)	Total (n = 51)			
IAPV	25 (83.3%)	1 (4.8%)	26 (51.0%)	96.1	83.3	95.2
KBV	30 (100%)	16 (76.2%)	46 (90.2%)	65.2	100	23.8
N. apis	27 (90%)	10 (47.6%)	37 (72.5%)	73.0	90.0	52.4
N. ceranae	30 (100%)	17 (80.9%)	47 (92.1%)	63.8	100	19.0
All 4 agents	23 (76.7%)	0 (0%)	23 (45.0%)	100	76.7	100

Sciencexpress

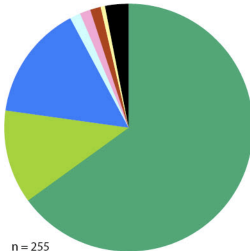
Fig. 1. Summary of bacterial groups from *A. mellifera* derived from colonies categorized as non-CCD and CCD. For both categories, the top blast hit for over 96% of sequences from 16S rRNA clones was a sequence obtained in previous studies on bacterial associates of *A. mellifera*. *The bacteria were categorized according to the cluster designations of Babendreier *et al*. (2007) [except for the Bifidobacterium-like sequence of Jeyaprakash *et al*. (2003)]. GenBank accession numbers corresponding to the categories are: γ-1 (AY370191-AY370192, DQ837602-DQ837609), γ-2 (DQ837610- DQ837611), β-1 (AY370189-AY370190, DQ837616- DQ837621), α-1 (AY370185-AY370187, DQ837622- DQ837624), α-2 (AY370188, DQ837625-DQ837626), F-4 (DQ837632- DQ837633), F-5 (AY370183, DQ837634- DQ837637), near Bifidobacterium (AY370184).

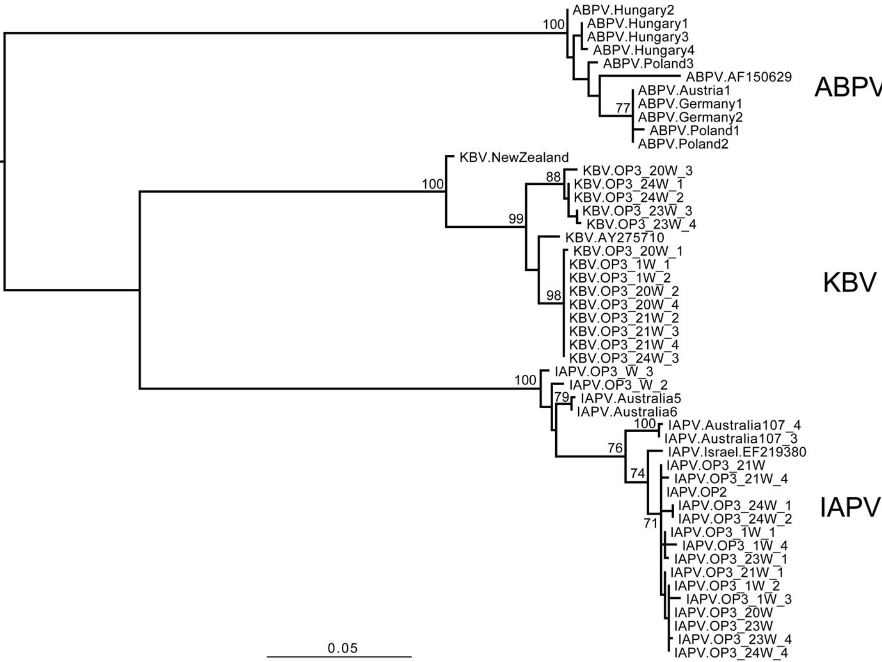
Fig. 2. Maximum likelihood phylogenetic tree of dicistroviruses based on a 741 nt sequence including the intergenic region (IGR). All branches are drawn to a scale of substitutions per site and bootstrap values are shown for key nodes. ABPV, Acute bee paralysis virus; IAPV, Israeli acute paralysis virus; KBV, Kashmir bee virus. All KBV sequences, except KBV.NewZealand and KBV.AY275710, were recovered in this study. All IAPV sequences, except IAPV.Israel.EF219380, were recovered in this study. GenBank accession numbers for the nucleotide sequences in file IAPVaug07IGR.sqn: IAPV.Australia107_4 EU122346, IAPV.Australia5 EU122347, IAPV.Australia6 EU122348, IAPV.Australia107_3 EU122349, IAPV.OP3_1W_1 EU122350, IAPV.OP3_1W_2 EU122351, IAPV.OP3_1W_3 EU122352, IAPV.OP3_1W_4 EU122353, IAPV.OP3_20W EU122354, IAPV.OP3_21W EU122355, IAPV.OP3_21W_1 EU122356, IAPV.OP3_W_2 EU122357, IAPV.OP3_W_3 EU122358, IAPV.OP3_21W_4 EU122359, IAPV.OP3_23W EU122360, IAPV.OP3_23W_1 EU122361, IAPV.OP3_23W_4 EU122362, IAPV.OP3_24W_1 EU122363, IAPV.OP3_24W_2 EU122364, IAPV.OP3_24W_4 EU122365, IAPV.OP2 EU122366, KBV.OP3_20W_1 EU122367, KBV.OP3_1W_1 EU122368 KBV.OP3_1W_2 EU122369, KBV.OP3_20W_2 EU122370, KBV.OP3_20W_3 EU122371, KBV.OP3_20W_4 EU122372, KBV.OP3_21W_2 EU122373, KBV.OP3_21W_3 EU122374, KBV.OP3_21W_4 EU122375, KBV.OP3_23W_3 EU122376, KBV.OP3_23W_4 EU122377, KBV.OP3_24W_1 EU122378, KBV.OP3_24W_2 EU122379, KBV.OP3_24W_3 EU122380.

Non-CCD colonies



CCD colonies




ABPV.Hungary2
ABPV.Hungary1
ABPV.Hungary3
ABPV.Hungary4
ABPV.Poland3
ABPV.AF150629
ABPV.Austria1
ABPV.Germany1
ABPV.Germany2
ABPV.Poland1
ABPV.Poland2
ABPV
KBV.NewZealand
KBV.OP3_20W_3
KBV.OP3_24W_1
KBV.OP3_24W_2
KBV.OP3_23W_3
KBV.OP3_23W_4
KBV.AY275710
KBV.OP3_20W_1
KBV.OP3_1W_1
KBV.OP3_1W_2
KBV.OP3_20W_2
KBV.OP3_20W_4
KBV.OP3_21W_2
KBV.OP3_21W_3
KBV.OP3_21W_4
KBV.OP3_24W_3
KBV
IAPV.OP3_W_3
IAPV.OP3_W_2
IAPV.Australia5
IAPV.Australia6
IAPV.Australia107_4
IAPV.Australia107_3
IAPV.Israel.EF219380
IAPV.OP3_21W
IAPV.OP3_21W_4
IAPV.OP2
IAPV.OP3_24W_1
IAPV.OP3_24W_2
IAPV.OP3_1W_1
IAPV.OP3_1W_4
IAPV.OP3_23W_1
IAPV.OP3_21W_1
IAPV.OP3_1W_2
IAPV.OP3_1W_3
IAPV.OP3_20W
IAPV.OP3_23W
IAPV.OP3_23W_4
IAPV.OP3_24W_4
IAPV
100
77
100
88
99
98
100
79
100
76
74
71
0.05