Data in Brief 36 (2021) 107019

Contents lists available at ScienceDirect

Data in Brief

journal homepage: www.elsevier.com/locate/dib

Data Article

Dataset of the next-generation sequencing of variable 16S rRNA from bacteria and ITS2 regions from fungi and plants derived from honeybees kept under anthropogenic landscapes

Marek Gancarz[a,b], Paul J. Hurd[c], Przemyslaw Latoch[d,e], Andrew Polaszek[f], Joanna Michalska-Madej[g], Łukasz Grochowalski[g], Dominik Strapagiel[g], Sebastian Gnat[h], Daniel Załuski[i], Robert Rusinek[a], Agata L. Starosta[e,j], Patcharin Krutmuang[k,l], Raquel Martín Hernández[m,n], Mariano Higes Pascual[m], Aneta A. Ptaszyńska[c,o,1,*]

[a] *Institute of Agrophysics, Polish Academy of Sciences, Doświadczalna 4 Str., 20-290 Lublin, Poland*
[b] *Faculty of Production and Power Engineering, University of Agriculture in Kraków, Balicka 116B, 30-149 Kraków, Poland*
[c] *School of Biological and Chemical Sciences, Queen Mary University of London, London E1 4NS, United Kingdom*
[d] *Polish-Japanese Academy of Information Technology, Koszykowa 86 st., 02-008 Warsaw, Poland*
[e] *Laboratory of Gene Expression, ECOTECH-Complex, Maria Curie-Sklodowska University, ul. Gleboka 39, 20-612 Lublin, Poland*
[f] *Department of Life Sciences, Insects Division, Natural History Museum, London SW7 5BD United Kingdom*
[g] *Biobank Lab, Department of Molecular Biophysics, Faculty of Biology and Environmental Protection, University of Łódź, Pilarskiego 14/16, 90-231 Łódź, Poland*
[h] *Faculty of Veterinary Medicine, Department of Veterinary Microbiology, Institute of Preclinical Veterinary Sciences, University of Life Sciences, Akademicka 12, 20-033 Lublin, Poland*
[i] *Department of Pharmaceutical Botany and Pharmacognosy, Ludwik Rydygier Collegium Medicum, Nicolaus Copernicus University, Marie Curie-Skłodowska 9, 85-094 Bydgoszcz, Poland*
[j] *Department of Molecular Biology, Institute of Biological Sciences, Maria Curie-Sklodowska University, Akademicka 19 Str., 20-033 Lublin, Poland*
[k] *Department of Entomology and Plant Pathology, Faculty of Agriculture, Chiang Mai University, 50200, Thailand*
[l] *Research Center of Microbial Diversity and Sustainable Utilization, Faculty of Science, Chiang Mai University, Chiang Mai 50200, Thailand*

* Corresponding author at: Department of Immunobiology, Institute of Biological Sciences, Faculty of Biology and Biotechnology, Maria Curie-Skłodowska University, Akademicka 19 Str., 20-033 Lublin, Poland.
E-mail address: aneta.ptaszynska@poczta.umcs.lublin.pl (A.A. Ptaszyńska).
[1] Senior author.

https://doi.org/10.1016/j.dib.2021.107019

[m] *IRIAF, Instituto Regional de Investigación y Desarrollo Agroalimentario y Forestal, Laboratorio de Patología Apícola, Centro de Investigación Apícola y Agroambiental (CIAPA), Consejería de Agricultura de la Junta de Comunidades de Castilla-La Mancha, Camino de San Martín s/n, 19180 Marchamalo, Spain*
[n] *Instituto de Recursos Humanos para la Ciencia y la Tecnología (INCRECYT-FEDER), Fundación Parque Científico y Tecnológico de Castilla—La Mancha, 02006 Albacete, Spain*
[o] *Department of Immunobiology, Institute of Biological Sciences, Faculty of Biology and Biotechnology, Maria Curie-Skłodowska University, Akademicka 19 Str., 20-033 Lublin, Poland*

A R T I C L E I N F O

Article history:
Received 7 March 2021
Revised 22 March 2021
Accepted 26 March 2021
Available online 2 April 2021

Keywords:
NGS, *Apis cerana*
Nosema sp.
Acarapis woodi
Trypanosomatida
Crithidia spp.
neogregarines
Apicystis spp.
Anthropocene

A B S T R A C T

Forager *Apis mellifera* honeybees were collected from four localities located in Europe, i.e.: London, UK; Athens, Greece; Marchamalo, Spain and Lublin, Poland. Furthermore, from Asia we have collected *A. mellifera* as well as *A. cerana* foragers form Chiang Mai in Thailand

We used next generation sequencing (NGS) to analyse the 16S rRNA bacterial gene amplicons based on the V3-V4 region and the ITS2 region from fungi and plants derived from honeybee samples. Amplicon libraries, were prepared using the 16S Metagenomic Sequencing Library Preparation, Preparing 16S Ribosomal RNA Gene Amplicons for the Illumina MiSeq System (Illumina®) protocol. NGS raw data are available at https://www.ncbi.nlm.nih.gov/bioproject/PRJNA686953. Furthermore, isolated DNA was used as the template for screening pathogens: *Nosema apis, N. ceranae, N. bombi*, tracheal mite (*Acarapis woodi*), any organism in the parasitic order Trypanosomatida, including *Crithidia* spp. (i.e., *Crithidia mellificae*), neogregarines including *Mattesia* and *Apicystis* spp. (i.e., *Apicistis bombi*).

The presented data can be used to compare the metagenomic samples from different honeybee population all over the world. A higher load of fungi, and bacteria groups such as: Firmicutes (*Lactobacillus*); ***γ***-proteobacteria, Neisseriaceae, and other unidentified bacteria was observed for *Nosema cearana* and neogregarines infected honeybees. Healthy honeybees had a higher load of plant pollens, and bacteria groups such as: *Orbales, Gilliamella, Snodgrassella,* and Enterobacteriaceae. More details can be found in research article [1] Ptaszyńska et al. 2021.

Specifications Table

Subject	Biological sciences: Entomology and insect science
Specific subject area	**Datasets:** Dataset 1. In the excel file are the raw original information form NGS of composition of bacteria from 16S_taxonomyReads from seasonal changes of Polish honeybee samples (collected from April to September). Dataset 2. In the excel file are the raw original information form NGS of composition of fungi and plant pollen from ITS_taxonomyReads from seasonal changes of Polish honeybee samples (collected from April to September).

(*continued on next page*)

	Dataset 3. In the excel file are the raw original information form NGS of composition of bacteria from 16S_taxonomyReads from UK, Greece, Spain and Thailand honeybee samples. Dataset 4. In the excel file are the raw original information form NGS of composition of bacteria from ITS_taxonomyReads from UK, Greece, Spain and Thailand honeybee samples. B. One-way ANOVA report from the correlation between UK, Spain, Greece and Thailand honeybees' health status and the detected fungi and plant pollens detected on the basis of ITS NGS analysis. Dataset 5. Table 1A shows the one-way ANOVA report from the correlation between Polish honeybees' health status and the bacteria detected on the base of 16S rDNA NGS analyses. In English. Dataset 6. Table 1B shows the one-way ANOVA report from the correlation between Polish honeybees' health status and fungi and plant pollens detected on the basis of ITS NGS analysis. In English. Dataset 7. Table 2A shows the one-way ANOVA report from the correlation between UK, Spain, Greece and Thailand honeybees' health status and the bacteria detected on the base of 16S rDNA NGS analyses. In English. Dataset 8. Table 2B shows the one-way ANOVA report from the correlation between UK, Spain, Greece and Thailand honeybees' health status and the detected fungi and plant pollens detected on the basis of ITS NGS analysis. In English. Dataset 9. Table 1A shows the one-way ANOVA report from the correlation between Polish honeybees' health status and the bacteria detected on the base of 16S rDNA NGS analyses. In Polish. Dataset 10. Table 1B shows the one-way ANOVA report from the correlation between Polish honeybees' health status and fungi and plant pollens detected on the basis of ITS NGS analysis. In Polish. Dataset 11. Table 2A shows the one-way ANOVA report from the correlation between UK, Spain, Greece and Thailand honeybees' health status and the bacteria detected on the base of 16S rDNA NGS analyses. In Polish. Dataset 12. Table 2B shows the one-way ANOVA report from the correlation between UK, Spain, Greece and Thailand honeybees' health status and the detected fungi and plant pollens detected on the basis of ITS NGS analysis. In Polish.
Type of data	Tables Figures
How data were acquired	NGS sequencing and the analysis of the 16S rRNA bacterial gene amplicon was based on the V3-V4 region and the ITS2 eukaryotic region for bee DNA samples. PL1-PL6 samples of *Apis mellifera* worker honeybees were collected from an urban apiary located in Lublin city, Poland, from April till September: PL1 (April), PL2 (May), PL3 (June), PL4 (July), PL5 (August), PL6 (September). Other samples of *Apis mellifera* worker honeybees were collected in July in London, UK (UK1, UK2); in November in Athens, Greece (GR1, GR2); in November in Marchamalo, Spain (ES1, ES2), and Chiang Mai, Thailand (TAI1, TAI2). From Chiang Mai, Thailand were also sampled *Apis cerana* worker honeybees (TAI3, TAI4). From one time and location, 3 specimens (forager honeybees) were taken, as the representative and consistent number for each group (data adequacy confirmed by the PCA analysis). Genomic DNA was extracted from whole honeybees using QIAamp DNA Kit according to manufacturer's instructions. Before pooling samples for libraries, the concentration was measured and the final concentration of pooled libraries for sequencing was 8 pM. Prepared libraries were sequenced on an Illumina MiSeq platform, 2 × 300 sequence reading in paired ends mode. The run contained PhiX libraries (PhiX Control Kit v3, Illumina®), to serve as an internal positive quality control. Amplicons for the 16S region and ITS2 were sequenced using the Illumina MiSeq platform. Data were trimmed and merged. For 16S analyses only full-length reads over 229 bp with medium length of all sequences at 414 bp were used [Table 1]. Sequences were assigned to taxonomy using classifier trained on SILVA 132 database with minimum similarity 90% of read matching to the reference. For ITS2 analyses only full-length reads over 269 bp with medium length of all sequences at 337 bp were used [Table 1].

(*continued on next page*)

	Sequences were assigned to taxonomy using classifier trained on all eukaryotes UNITE database v8.2 with the minimum similarity of 90% of the read matching to the reference [2,3]. Obtained data alpha rarefaction by country is shown on Fig. 1 and Principal Component Analysis (PCA), using Jaccard 's similarity based on taxonomy composition of 16S rRNA and ITS2 on Fig. 2. Isolated DNA was used as the template for screening pathogens: *Nosema apis, Nosema ceranae, Nosema bombi*, tracheal mite (*Acarapis woodi*), any organism in the parasitic order Trypanosomatida, including *Crithidia* spp. (i.e., *Crithidia mellificae*), neogregarines including *Mattesia* and *Apicystis* spp. (i.e., *Apicistis bombi*). The presence of pathogens [Table 3] in collected bee samples was detected using ITS2 amplicon data and specific primers under standard PCR, according to methodology described to *Nosema apis* by Martín-Hernández et al. [4], *Nosema ceranae* by Martín-Hernández et al. [4]; *Nosema bombi* by Klee et al. [5] Tracheal mite (*Acarapis woodi*) by Yang et al. [6]; any organism in the parasitic order Trypanosomatida, including *Crithidia* spp. (i.e. *Crithidia mellificae*) Meeus et al. [6]; neogregarines including *Mattesia* and *Apicystis* spp. (i.e. *Apicistis bombi*) Meeus et al. [6].
Data format	Raw Filtered
Parameters for data collection	Forager honeybees were recognized as bees returning to the hive and captured at the hive entrance around midday. Genomic DNA was extracted from whole honeybees using QIAamp DNA Kit according to manufacturer's instructions. Isolates were sent to the Biobank, Poland for NGS analysis.
Description of data collection	Forager honeybees were recognized as bees returning to the hive and captured at the hive entrance around midday. All foragers were captured individually, using tweezers.
Data source location	Data source locations are presented in Table 2.
Data accessibility	All raw data are available at https://www.ncbi.nlm.nih.gov/bioproject/PRJNA686953 Datasets are available at: Ptaszynska, Aneta A (2021), "Dataset of the next-generation sequencing of variable 16S rRNA from bacteria and ITS2 regions from fungi and plants derived from honeybees kept under anthropogenic landscapes", Mendeley Data, V1, http://dx.doi.org/10.17632/5zrz4fmw5y.2
Related research article	[1] Ptaszyńska A.A., Latoch P., Hurd P.J., Polaszek A., Michalska-Madej J., Grochowalski Ł., Strapagiel D., Gnat S., Załuski D., Gancarz M., Rusinek R., Krutmuang P., Martín Hernández R., Higes Pascual M., Starosta A.L. 2021. Amplicon sequencing of variable 16S rRNA from bacteria, and ITS2 regions from fungi and plants, reveals honeybee susceptibility to diseases results from their forage availability under anthropogenic landscapes. MDPI Pathogens. 10,3 http://dx.doi.org/10.3390/pathogens10030381

Value of the Data

- Next-generation sequencing (NGS) has revolutionized the biological sciences and obtained data can help analysing bee biology, food preferences and susceptibility to diseases.
- Standardized data collection of honeybee microbiome derived from NGS data is crucial for proper data analysis.
- Urban beekeeping is under urgent studies due to pollinator crisis and honeybee NGS data can be useful to construct an urban ecological network.
- Correlation of honeybee microbiome from NGS data with pathogens can lead to new forms of active protection of pollinators.
- High loads of bacteria such as: *Orbales, Gilliamella, Snodgrassella,* Enterobacteriaceae and plant pollen can be used as honeybee well-being indicators.

1. Data Description

1.1. Datasets

NGS raw data are available at https://www.ncbi.nlm.nih.gov/bioproject/PRJNA686953.

Datasets 1-12 are available at: Ptaszynska, Aneta A (2021), "Dataset of the next-generation sequencing of variable 16S rRNA from bacteria and ITS2 regions from fungi and plants derived

from honeybees kept under anthropogenic landscapes", Mendeley Data, V1, http://dx.doi.org/10.17632/5zrz4fmw5y.1 Table 1. describes sequences filtering statistics Input – initial number of sequences, Filtered – number of reads after removing low-quality data, Denoised – number of reads after removing data considered as noise, Merged – number of correctly merged forward and reverse reads, Non-chimeric – number of sequences after chimera removal; final number of reads. Table 1a. describes 16S reads from Polish samples. *Apis mellifera* worker honeybees sampled from April till September: PL1 (April), PL2 (May), PL3 (June), PL4 (July), PL5 (August), PL6 (September). Honeybees were collected from an urban apiary located in Lublin city, Poland. Table 1b. describes ITS2 reads from Polish samples. *Apis mellifera* worker honeybees sampled from April till September: PL1 (April), PL2 (May), PL3 (June), PL4 (July), PL5 (August), PL6 (September). Honeybees were collected from an urban apiary located in Lublin city, Poland. Table 1c. describes 16Sreads from UK, GR, ES, TAI samples. *Apis mellifera* worker honeybees sampled in July in London, UK (UK1, UK2); in November in Athens, Greece (GR1, GR2); in November in Marchamalo, Spain (ES1, ES2), and Chiang Mai, Thailand (TAI1, TAI2). From Chiang Mai, Thailand were also sampled *Apis cerana* worker honeybees (TAI3, TAI4). Table 1d. describes ITS2 reads from UK, GR, ES, TAI samples. *Apis mellifera* worker honeybees sampled in July in London, UK (UK1, UK2); in November in Athens, Greece (GR1, GR2); in November in Marchamalo, Spain (ES1, ES2), and Chiang Mai, Thailand (TAI1, TAI2). From Chiang Mai, Thailand were also sampled *Apis cerana* worker honeybees (TAI3, TAI4). Table 2. describes localities of investigated samples.

Ptaszynska, Aneta A (2021), "Dataset of the next-generation sequencing of variable 16S rRNA from bacteria and ITS2 regions from fungi and plants derived from honeybees kept under anthropogenic landscapes", Mendeley Data, V1, http://dx.doi.org/10.17632/5zrz4fmw5y.2.

Dataset 1. Excel 1_16S_taxonomyReads_BEES-PL. In the excel file are the row original information form NGS of composition of bacteria from 16S_taxonomyReads from seasonal changes of Polish honeybee samples (collected from April to September).

Apis mellifera worker honeybees sampled from April till September: PL1 (April), PL2 (May), PL3 (June), PL4 (July), PL5 (August), PL6 (September). Honeybees were collected from an urban apiary located in Lublin city, Poland.

Description of the taxonomy Excel tables

The excel tables contain the composition of 7 taxonomic level for the individual samples. The sampling depth for 16S sequencing data is 14096 and for ITS data 34100. This is the total sum of the reads at each level.

Levels denotes accordingly:
Level 1 = Kingdom,
Level 2 = Phylum,
Level 3 = Class,
Level 4 = Order,
Level 5 = Family,
Level 6 = Genus,
Level 7 = Species.

Dataset 2. Excel 2_ITS_taxonomyReads_BEES-PL. In the excel file are the row original information form NGS of composition of fungi and plant pollen from ITS_taxonomyReads from seasonal changes of Polish honeybee samples (collected from April to September).

Description of the taxonomy Excel tables

The excel tables contain the composition of 7 taxonomic level for the individual samples. The sampling depth for 16S sequencing data is 14096 and for ITS data 34100. This is the total sum of the reads at each level.

Dataset 3. Excel 3_16S_taxonomyReads_BEES-other. In the excel file are the row original information form NGS of composition of bacteria from 16S_taxonomyReads from UK, Greece, Spain and Thailand honeybee samples.

Apis mellifera worker honeybees sampled in July in London, UK (UK1, UK2); in November in Athens, Greece (GR1, GR2); in November in Marchamalo, Spain (ES1, ES2), and Chiang Mai, Thailand (TAI1, TAI2). From Chiang Mai, Thailand were also sampled *Apis cerana* worker honeybees (TAI3, TAI4).

Table 1
Sequences filtering statistics.

- Input – initial number of sequences,
- Filtered – number of reads after removing low-quality data,
- Denoised – number of reads after removing data considered as noise,
- Merged – number of correctly merged forward and reverse reads,
- Non-chimeric – number of sequences after chimera removal; final number of reads.

16S – PL.
Apis mellifera worker honeybees sampled from April till September: PL1 (April), PL2 (May), PL3 (June), PL4 (July), PL5 (August), PL6 (September). Honeybees were collected from an urban apiary located in Lublin city, Poland.

sample-id	input	filtered	percentage of input passed filter	denoised	merged	percentage of input merged	non-chimeric	percentage of input non-chimeric
PL1(1)	38239	20701	54,14	20637	20406	53,36	19902	52,05
PL1(2)	44576	29660	66,54	29562	29288	65,7	28528	64
PL1(3)	49083	26149	53,28	26064	25771	52,5	25092	51,12
PL2(1)	40017	24250	60,6	24216	24095	60,21	23239	58,07
PL2(2)	62869	39145	62,26	39058	38865	61,82	37217	59,2
PL3(2)	22213	14178	63,83	14140	14114	63,54	14096	63,46
PL3(3)	41999	31808	75,74	31750	31707	75,49	31521	75,05
PL4(1)	61900	41059	66,33	40969	40507	65,44	38764	62,62
PL4(2)	54334	36764	67,66	36575	36270	66,75	34761	63,98
PL4(3)	72110	49176	68,2	49052	48547	67,32	46512	64,5
PL5(1)	46551	28087	60,34	27929	27743	59,6	27473	59,02
PL5(2)	41235	25786	62,53	25673	25551	61,96	25359	61,5
PL5(3)	43505	28996	66,65	28945	28898	66,42	28820	66,25
PL6(1)	44911	24875	55,39	24748	24401	54,33	22598	50,32
PL6(2)	37642	20884	55,48	20784	20462	54,36	18948	50,34
PL6(3)	63812	34485	54,04	34327	33869	53,08	31062	48,68

ITS2 – PL.
Apis mellifera worker honeybees sampled from April till September: PL1 (April), PL2 (May), PL3 (June), PL4 (July), PL5 (August), PL6 (September). Honeybees were collected from an urban apiary located in Lublin city, Poland.

sample-id	input	filtered	percentage of input passed filter	denoised	merged	percentage of input merged	non-chimeric	percentage of input non-chimeric
PL1(1)	160044	51024	31,88	50799	44057	27,53	43528	27,2
PL1(2)	147968	39779	26,88	39609	34607	23,39	34223	23,13
PL1(3)	166321	45521	27,37	45319	39417	23,7	38953	23,42
PL2(1)	161433	123703	76,63	123528	122002	75,57	113293	70,18
PL2(2)	120890	90633	74,97	90481	89267	73,84	82117	67,93
PL2(3)	140416	102813	73,22	102632	102355	72,89	100746	71,75
PL3(1)	147599	106749	72,32	106540	105072	71,19	102410	69,38
PL3(2)	125356	94618	75,48	94229	92777	74,01	90241	71,99
PL3(3)	91520	64349	70,31	64054	62253	68,02	60547	66,16
PL4(1)	157768	84762	53,73	84292	82715	52,43	81679	51,77
PL4(2)	129401	79102	61,13	78655	76783	59,34	75749	58,54
PL4(3)	113704	66404	58,4	66103	64776	56,97	63820	56,13
PL5(1)	170603	128436	75,28	128035	126164	73,95	123249	72,24
PL5(2)	162549	120393	74,07	120123	118010	72,6	115383	70,98
PL5(3)	144894	98781	68,17	98569	96751	66,77	94714	65,37
PL6(1)	148006	108317	73,18	107808	106726	72,11	105435	71,24
PL6(2)	141798	100218	70,68	100005	99095	69,88	97342	68,65
PL6(3)	130099	95359	73,3	95236	94250	72,44	93123	71,58

16S –UK, GR, ES, TAI.
Apis mellifera worker honeybees sampled in July in London, UK (UK1, UK2); in November in Athens, Greece (GR1, GR2); in November in Marchamalo, Spain (ES1, ES2), and Chiang Mai, Thailand (TAI1, TAI2). From Chiang Mai, Thailand were also sampled *Apis cerana* worker honeybees (TAI3, TAI4).

sample-id	input	filtered	percentage of input passed filter	denoised	merged	percentage of input merged	non-chimeric	percentage of input non-chimeric
UK-1	125453	70721	56,37%	70261	69149	55,12%	62612	49,91%
UK-2	113210	60764	53,67%	60625	60386	53,34%	58239	51,44%
GR-1	189774	101189	53,32%	100943	98753	52,04%	98281	51,79%
GR-2	219902	124908	56,80%	124282	122054	55,50%	107053	48,68%
ES-1	165148	91760	55,56%	91535	91274	55,27%	89325	54,09%
ES-2	135182	73017	54,01%	72836	72647	53,74%	72582	53,69%
TAI-1	205607	122277	59,47%	122053	121803	59,24%	120289	58,50%
TAI-2	275928	158746	57,53%	158498	157855	57,21%	156277	56,64%
TAI-3	247489	148904	60,17%	148390	147315	59,52%	136300	55,07%
TAI-4	233312	137798	59,06%	137420	136713	58,60%	132126	56,63%

(*continued on next page*)

Table 1 (*continued*)

ITS2 – UK, GR, ES, TAI.
Apis mellifera worker honeybees sampled in July in London, UK (UK1, UK2); in November in Athens, Greece (GR1, GR2); in November in Marchamalo, Spain (ES1, ES2), and Chiang Mai, Thailand (TAI1, TAI2). From Chiang Mai, Thailand were also sampled *Apis cerana* worker honeybees (TAI3, TAI4).

sample-id	input	filtered	percentage of input passed filter	denoised	merged	percentage of input merged	non-chimeric	percentage of input non-chimeric
UK-1	60154	35292	58,67%	35052	34656	57,61%	34132	56,74%
UK-2	141920	76895	54,18%	76768	74365	52,40%	74196	52,28%
GR-1	192913	83547	43,31%	83462	82158	42,59%	81601	42,30%
GR-2	121281	67887	55,97%	67627	64625	53,29%	61797	50,95%
ES-1	263444	112349	42,65%	112231	105442	40,02%	105254	39,95%
ES-2	152574	108152	70,88%	108133	107913	70,73%	107913	70,73%
TAI-1	297119	148360	49,93%	148075	121757	40,98%	121337	40,84%
TAI-2	254112	138951	54,68%	138582	136814	53,84%	135843	53,46%
TAI-3	321713	160884	50,01%	160527	133804	41,59%	130792	40,65%
TAI-4	206686	123427	59,72%	123298	121189	58,63%	119821	57,97%

Table 2
Localities of investigated samples.

Country	City	Geographical coordinates	Sample abbreviation	Time of samplings	Organisms
Poland	Lublin	51°15′N 22°34′E	PL1	April	*Apis melliefera*
			PL2	May	*Apis melliefera*
			PL3	June	*Apis melliefera*
			PL4	July	*Apis melliefera*
			PL5	August	*Apis melliefera*
			PL6	September	*Apis melliefera*
UK	London	51°52′N 0°03′W	UK1	July	*Apis melliefera*
		51°29′N 0°10′W	UK2	July	*Apis melliefera*
Greece	Athens	37°59′N 23°42′E	GR1	November	*Apis melliefera*
			GR2	November	*Apis melliefera*
Spain	Marchamalo	40°68′N 3°21′W	ES1	November	*Apis melliefera*
			ES2	November	*Apis melliefera*
Thailand	Chiang Mai	18°50′ 98°58″E	TAI1	February	*Apis melliefera*
			TAI2	February	*Apis melliefera*
			TAI3	February	*Apis cerana*
			TAI4	February	*Apis cerana*

Description of the taxonomy Excel tables

The excel tables contain the composition of 7 taxonomic level for the individual samples. The sampling depth for 16S sequencing data is 14096 and for ITS data 34100. This is the total sum of the reads at each level.

Levels denotes accordingly:
Level 1 = Kingdom,
Level 2 = Phylum,
Level 3 = Class,
Level 4 = Order,
Level 5 = Family,
Level 6 = Genus,
Level 7 = Species.

Dataset 4. Excel 4_ITS_taxonomyReads_BEES-other. In the excel file are the row original information form NGS of composition of bacteria from ITS_taxonomyReads from UK, Greece, Spain and Thailand honeybee samples.

Description of the taxonomy Excel tables

The excel tables contain the composition of 7 taxonomic level for the individual samples. The sampling depth for 16S sequencing data is 14096 and for ITS data 34100. This is the total sum of the reads at each level.

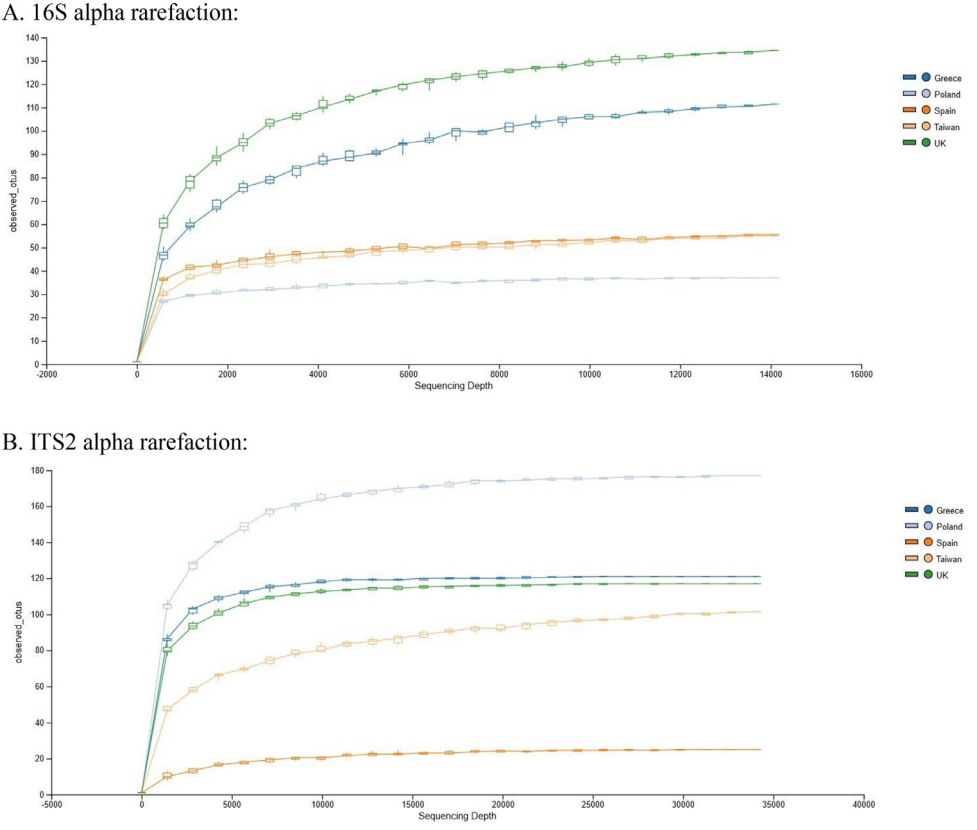

Fig. 1. Shows the alpha-rarefaction by country. The sampling depth was set at 14096 for 16S amplicon sequencing data and 34100 for ITS amplicon data. This parameter was selected to include all available samples in the analysis and, as can be seen in the graphs above, it is sufficient to show the full taxonomic diversity for samples from individual countries. *Apis mellifera* worker honeybees sampled in July in London, UK (UK1, UK2); in November in Athens, Greece (GR1, GR2); in November in Marchamalo, Spain (ES1, ES2), and Chiang Mai, Thailand (TAI1, TAI2). From Chiang Mai, Thailand were also sampled *Apis cerana* worker honeybees (TAI3, TAI4).1A. shows the 16S alpha-rarefaction by country1B. shows ITS2 alpha rarefaction by country.

Dataset 5. The table 1A reports the sequences filtering statistics of one-way ANOVA from the correlation between Polish honeybees' health status and the bacteria detected on the base of 16S rDNA NGS analyses. For the ANOVA test, the level of statistical significance was assumed to be $\alpha = 0.05$ and the same level of statistical significance was used in all comparisons. The results for which p values equal to or less than 0.05 were obtained differ significantly from each other. *Apis mellifera* worker honeybees sampled from April till September: PL1 (April), PL2 (May), PL3 (June), PL4 (July), PL5 (August), PL6 (September). Honeybees were collected from an urban apiary located in Lublin city, Poland.

Dataset 6. The table 1B reports the sequences filtering statistics of one-way ANOVA from the correlation between Polish honeybees' health status and fungi and plant pollens detected on the basis of ITS NGS analysis. For the ANOVA test, the level of statistical significance was assumed to be $\alpha = 0.05$ and the same level of statistical significance was used in all comparisons. The results for which p values equal to or less than 0.05 were obtained differ significantly from each other. *Apis mellifera* worker honeybees sampled from April till September: PL1 (April), PL2 (May), PL3 (June), PL4 (July), PL5 (August), PL6 (September). Honeybees were collected from an urban apiary located in Lublin city, Poland.

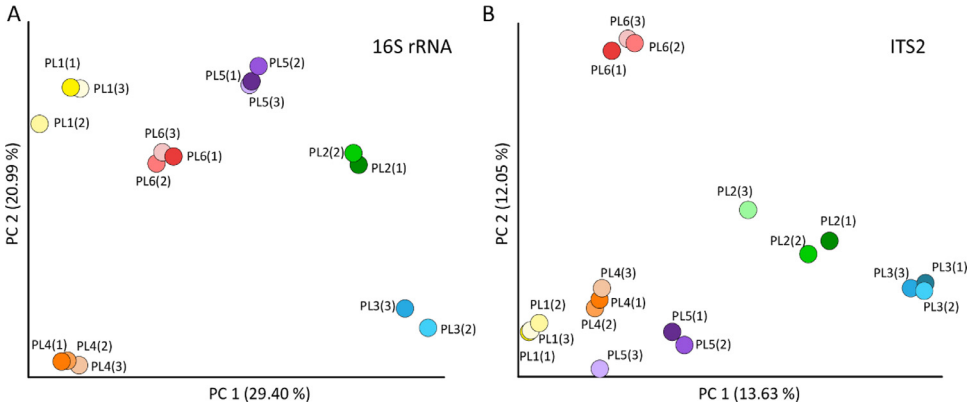

Fig. 2. Shows Principal Component Analysis (PCA), using Jaccard 's similarity based on taxonomy composition of 16S rRNA and ITS2. *Apis mellifera* worker honeybees sampled from April till September: PL1 (April), PL2 (May), PL3 (June), PL4 (July), PL5 (August), PL6 (September). Honeybees were collected from an urban apiary located in Lublin city, Poland.

Dataset 7. The table 2A reports the sequences filtering statistics of one-way ANOVA from the correlation between UK, Spain, Greece and Thailand honeybees' health status and the bacteria detected on the base of 16S rDNA NGS analyses. For the ANOVA test, the level of statistical significance was assumed to be $\alpha = 0.05$ and the same level of statistical significance was used in all comparisons. The results for which p values equal to or less than 0.05 were obtained differ significantly from each other. *Apis mellifera* worker honeybees sampled in July in London, UK (UK1, UK2); in November in Athens, Greece (GR1, GR2); in November in Marchamalo, Spain (ES1, ES2), and Chiang Mai, Thailand (TAI1, TAI2). From Chiang Mai, Thailand were also sampled *Apis cerana* worker honeybees (TAI3, TAI4).

Dataset 8. The table 2B reports the sequences filtering statistics of one-way ANOVA from the correlation between UK, Spain, Greece and Thailand honeybees' health status and the detected fungi and plant pollens detected on the basis of ITS NGS analysis. For the ANOVA test, the level of statistical significance was assumed to be $\alpha = 0.05$ and the same level of statistical significance was used in all comparisons. The results for which p values equal to or less than 0.05 were obtained differ significantly from each other. *Apis mellifera* worker honeybees sampled in July in London, UK (UK1, UK2); in November in Athens, Greece (GR1, GR2); in November in Marchamalo, Spain (ES1, ES2), and Chiang Mai, Thailand (TAI1, TAI2). From Chiang Mai, Thailand were also sampled *Apis cerana* worker honeybees (TAI3, TAI4).

Dataset 9. The table 1A reports the sequences filtering statistics of one-way ANOVA report from the correlation between Polish honeybees' health status and the bacteria detected on the base of 16S rDNA NGS analyses. For the ANOVA test, the level of statistical significance was assumed to be $\alpha = 0.05$ and the same level of statistical significance was used in all comparisons. The results for which p values equal to or less than 0.05 were obtained differ significantly from each other. *Apis mellifera* worker honeybees sampled from April till September: PL1 (April), PL2 (May), PL3 (June), PL4 (July), PL5 (August), PL6 (September). Honeybees were collected from an urban apiary located in Lublin city, Poland. In Polish.

Dataset 10. The table 1B reports the sequences filtering statistics of one-way ANOVA report from the correlation between Polish honeybees' health status and fungi and plant pollens detected on the basis of ITS NGS analysis. For the ANOVA test, the level of statistical significance was assumed to be $\alpha = 0.05$ and the same level of statistical significance was used in all comparisons. The results for which p values equal to or less than 0.05 were obtained differ significantly from each other. *Apis mellifera* worker honeybees sampled from April till September: PL1 (April), PL2 (May), PL3 (June), PL4 (July), PL5 (August), PL6 (September). Honeybees were collected from an urban apiary located in Lublin city, Poland. In Polish.

Table 3
Describes the presence of pathogens in collected bee samples.

Sample abbreviation	Time of samplings	Presence of pathogens based on ITS2 and PCR detection* of: • *Nosema apis* • *N. ceranae* • *N. bombi* • tracheal mite (*Acarapis woodi*) • any organism in the parasitic order Trypanosomatida, including *Crithidia* spp. (i.e. *Crithidia mellificae*); • neogregarines including *Mattesia* and *Apicystis* spp. (i.e. *Apicistis bombi*).
PL1	April	• *Nosema ceranae*
PL2	May	–
PL3	June	• *Betsia* sp.
PL4	July	• *Nosema ceranae* • neogregarines
PL5	August	–
PL6	September	• *Nosema ceranae* • neogregarines
UK1	July	–
UK2	July	• neogregarines
GR1	November	• Cyanobacteria • *Nosema ceranae* • neogregarines
GR2	November	–
ES1	November	–
ES2	November	• *Nosema ceranae* • neogregarines
TAI1	February	• *Nosema ceranae* • neogregarines
TAI2	February	–
TAI3	February	–
TAI4	February	• *Nosema ceranae* • Neogregarines

* Pathogens detected using ITS2 amplicon data and specific primers under standard PCR, according to methodology described to *Nosema apis* by Martín-Hernández et al. [4], *Nosema ceranae* by Martín-Hernández et al. [4]; *Nosema bombi* by Klee et al. [5] Tracheal mite (*Acarapis woodi*) by Yang et al. [6]; any organism in the parasitic order Trypanosomatida, including *Crithidia* spp. (i.e. *Crithidia mellificae*) Meeus et al. [7]; neogregarines including *Mattesia* and *Apicystis* spp. (i.e. *Apicistis bombi*) Meeus et al. [7]; – no detected pathogens.

Dataset 11. The table 2A reports the sequences filtering statistics of one-way ANOVA report from the correlation between UK, Spain, Greece and Thailand honeybees' health status and the bacteria detected on the base of 16S rDNA NGS analyses. For the ANOVA test, the level of statistical significance was assumed to be $\alpha = 0.05$ and the same level of statistical significance was used in all comparisons. The results for which p values equal to or less than 0.05 were obtained differ significantly from each other. *Apis mellifera* worker honeybees sampled in July in London, UK (UK1, UK2); in November in Athens, Greece (GR1, GR2); in November in Marchamalo, Spain (ES1, ES2), and Chiang Mai, Thailand (TAI1, TAI2). From Chiang Mai, Thailand were also sampled *Apis cerana* worker honeybees (TAI3, TAI4). In Polish.

Dataset 12. The table 2B reports the sequences filtering statistics of one-way ANOVA report from the correlation between UK, Spain, Greece and Thailand honeybees' health status

and the detected fungi and plant pollens detected on the basis of ITS NGS analysis. For the ANOVA test, the level of statistical significance was assumed to be $\alpha = 0.05$ and the same level of statistical significance was used in all comparisons. The results for which p values equal to or less than 0.05 were obtained differ significantly from each other. *Apis mellifera* worker honeybees sampled in July in London, UK (UK1, UK2); in November in Athens, Greece (GR1, GR2); in November in Marchamalo, Spain (ES1, ES2), and Chiang Mai, Thailand (TAI1, TAI2). From Chiang Mai, Thailand were also sampled *Apis cerana* worker honeybees (TAI3, TAI4). In Polish.

2. Experimental Design, Materials and Methods

2.1. Materials and methods

2.1.1. Honeybee collection and DNA isolation

Forager honeybees were captured from five localities situated in urban areas of Poland, UK, Spain, Greece and Thailand (Table 2). Genomic DNA was extracted from whole honeybees using QIAamp DNA Kit according to manufacturer's instructions. Isolates were sent to the Biobank, Poland for NGS analysis.

Isolated DNA was used as the template for screening pathogens: *Nosema apis, Nosema ceranae, Nosema bombi*, tracheal mite (*Acarapis woodi*), any organism in the parasitic order Trypanosomatida, including Crithidia spp. (i.e., *Crithidia mellificae*), neogregarines including *Mattesia* and *Apicystis* spp. (i.e., *Apicistis bombi*), using PCR techniques described earlier [4-7]. Detected pathogens are listed in Table 3.

Ethics Statement

Although no permission is needed to administer experiments on insects, our research was planned in a way that reduced the number of honeybees to the minimum necessary for the proper conducting of these experiments.

CRediT Author Statement

Author Contributions: **Marek Gancarz**, and **Robert Rusinek:** analysed obtained data, interpreted the results, co-wrote the paper; **Paul J. Hurd:** analysed data, especially of metabiom and parasites, co-wrote the paper; **Przemyslaw Latoch:** analysed data. Patcharin Krutmuang, analysed Thai data, co-wrote the paper; **Raquel Martín Hernández**, and **Mariano Higes Pascual:** analysed UK data, co-wrote the paper; **Aneta A. Ptaszyńska:** co-wrote the paper; **Joanna Michalska-Madej:** conducted laboratory work for sequencing library preparation, sequencing, and detection of pathogens; **Łukasz Grochowalski:** analysed raw data from metabiom sequencing, prepared tables, co-wrote the paper; **Agata L. Starosta:** analysed data. Dominik Strapagiel, analysed data from metabiom sequencing, co-wrote the paper; **Sebastian Gnat:** performed genetic analyses; **Daniel Załuski:** drafted and made a correction of the manuscript; **Aneta A. Ptaszyńska:** (senior author), designed the experiments, analysed data, and wrote the paper.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships which have or could be perceived to have influenced the work reported in this article.

Acknowledgments

The publication of the article was financed by the Polish National Agency for Academic Exchange under the Foreign Promotion Programme (NAWA), for AAP (bee-research.umcs.pl; Api Lab UMCS PPI/PZA/2019/1/00039). Honeybees were collected during the Miniatura 2 project ID 418332 founded by the NCN for AAP and the first research plans were possible thanks to EU: GB-TAF-7137 SYNTHESYS project for AAP. Work in the Hurd lab was funded by the BBSRC (BB/L023164/1) and granted to PJH.

References

[1] A.A. Ptaszyńska, P. Latoch, P.J. Hurd, A. Polaszek, J. Michalska-Madej, Ł. Grochowalski, D. Strapagiel, S. Gnat, D. Załuski, M. Gancarz, R. Rusinek, P. Krutmuang, R. Martín Hernández, M. Higes Pascual, A.L. Starosta, Amplicon sequencing of variable 16S rRNA from bacteria, and ITS2 regions from fungi and plants, reveals honeybee suscep-tibility to diseases results from their forage availability under anthropogenic landscapes, MDPI Pathogens 10 (2021) 3, doi:10.3390/pathogens10030381.

[2] A Klindworth, E Pruesse, T Schweer, J Peplies, C Quast, M Horn, FO Glöckner, Evaluation of general 16S ribosomal RNA gene PCR primers for classical and next-generation sequencing-based diversity studies, Nucleic Acids Res. 41 (1) (2013 Jan 7) e1, doi:10.1093/nar/gks808.

[3] T.J. White, T.D. Bruns, S.B. Lee, J.W. Taylor, M.A. Innis, D.H. Gelfand, J. Sninsky, Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics, in: M.A. Innis, D.H. Gelfand,, J.J. Sninsky, T.J. White (Eds.), PCR Protocols: A Guide to Methods and Applications, Academic Press, San Diego, CA, 1990, pp. 315–322.

[4] R Martin-Hernandez, A Meana, L Prieto, AM Salvador, E Garrido-Bailon, M. Higes, Outcome of the colonization of Apis mellifera by Nosema ceranae, Appl. Environ. Microbiol. 73 (2007) 6331–6338.

[5] J Klee, W Tek Tay, RJ Paxton, Specific and sensitive detection of Nosema bombi (Microsporidia: Nosematidae) in bumble bees (Bombus spp.; Hymenoptera: Apidae) by PCR of partial rRNA gene sequences, J. Invertebr. Pathol. 91 (2) (2006 Feb) 98–104 Epub 2005 Dec 22, doi:10.1016/j.jip.2005.10.012.

[6] B Yang, G Peng, T Li, T Kadowaki, Molecular and phylogenetic characterization of honey bee viruses, Nosema microsporidia, protozoan parasites, and parasitic mites in China, Ecol. Evol. 3 (2) (2013) 298–311, doi:10.1002/ece3.464.

[7] I Meeus, DC de Graaf, K Jans, G Smagghe, Multiplex PCR detection of slowly-evolving trypanosomatids and neogregarines in bumblebees using broad-range primers, J. Appl. Microbiol. 109 (1) (2010 Jul) 107–115 Epub 2009 Nov 23, doi:10.1111/j.1365-2672.2009.04635.x.