Li *Environ Sci Eur* (2019) 31:37
https://doi.org/10.1186/s12302-019-0217-x

Environmental Sciences Europe

RESEARCH Open Access

Metagenomic screening of microbiomes identifies pathogen-enriched environments

Xiaofang Li*

Abstract

Background: Human pathogens are widespread in the environment, and examination of pathogen-enriched environments in a rapid and high-throughput fashion is important for development of pathogen-risk precautionary measures. In this study, a Local BLASTP procedure for metagenomic screening of pathogens in the environment was developed using a toxin-centered database. A total of 69 microbiomes derived from ocean water, freshwater, soils, feces, and wastewater were screened using the Local BLASTP procedure. Bioinformatic analysis and Canonical Correspondence Analysis were conducted to examine whether the toxins included in the database were taxonomically associated.

Results: The specificity of the Local BLASTP method was tested with known and unknown toxin sequences. Bioinformatic analysis indicated that most toxins were phylum-specific but not genus-specific. Canonical Correspondence Analysis implied that almost all of the toxins were associated with the phyla of *Proteobacteria*, *Nitrospirae* and *Firmicutes*. Local BLASTP screening of the global microbiomes showed that pore-forming RTX toxin, ornithine carbamoyltransferase ArgK, and RNA interferase Rel were most prevalent globally in terms of relative abundance, while polluted water and feces samples were the most pathogen-enriched.

Conclusions: The Local BLASTP procedure was applied for rapid detection of toxins in environmental samples using a toxin-centered database built in this study. Screening of global microbiomes in this study provided a quantitative estimate of the most prevalent toxins and most pathogen-enriched environments. Feces-contaminated environments are of particular concern for pathogen risks.

Keywords: Metagenomics, Microbiome, Local BLASTP, Toxins, Pathogens

Background

Rapid identification of pathogens in a particular environment is important for pathogen-risk management. Human pathogens are ubiquitous in the environment, and infections from particular environments have been reported worldwide. For example, soil-related infectious diseases are common [1, 2]. *Legionella longbeachae* infection has been reported in many cases, mainly due to potting mixes and composts [3]. Survival of enteric viruses and bacteria has also been detected in various water environments, including aquifers and lakes [4–7].

Examination of pathogens from infected individuals with a particular clinical syndrome has been a major achievement of modern medical microbiology [8]. Nevertheless, we still know little about the magnitude of the abundance and diversity of known common pathogens in various environments, which is very important for the development of appropriate precautions for individuals who come in contact with certain environmental substrates. This can be realized through metagenomic detection of pathogenic factors in a time-efficient and high-throughput manner using next-generation sequencing methods [8].

Metagenomic detection of pathogens can be accomplished through different schemes. Li et al. [9] examined the level and diversity of bacterial pathogens in sewage treatment plants using a 16S rRNA amplicon-based

*Correspondence: xfli@sjziam.ac.cn
Key Laboratory of Agricultural Water Resources, Hebei Key Laboratory of Soil Ecology, Centre for Agricultural Resources Research, Institute of Genetics and Developmental Biology, Chinese Academy of Sciences, No. 286, Huaizhong Rd, Shijiazhuang 050021, China

Springer Open

© The Author(s) 2019. This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

metagenomic procedure. Quantitative PCR has also been applied for monitoring specific pathogens in wastewater [10]. More studies have applied the whole-genome-assembly scheme to detect one or multiple dominant pathogens, most of which were for viral detection in clinical samples [11–14]. Although metagenomic-based whole-genome-assembly for bacterial pathogen detection can be conducted at the single species level [15], its computational requirements are high if it is in a high-throughput fashion. In 2014, Baldwin et al. [16] designed the PathoChip for screening pathogens in human tissues by targeting unique sequences of viral and prokaryotic genomes with multiple probes in a microarray. This approach can screen virtually all pathogen-enriched samples in a high-throughput manner.

Despite the aforementioned progress in metagenomic tools for pathogen detection, metagenomic screening for bacterial pathogens in environments such as soil, where microbial diversity is tremendous, is still challenging. This is mostly due to difficulty in assembling short reads generated by next-generation sequencing [8]. The whole-genome-assembly approach is efficient at identifying viromes, but not at dealing with bacterial pathogens from metagenomes especially when target pathogens are of low abundance. Amplicon-based approaches are able to detect bacterial pathogens in a high-throughput manner; however, it is well known that phenotypic diversity exists widely across and within microbial species of a genus because of divergent evolution [17, 18]. This also holds true for pathogenic factors [19]. Moreover, toxin factors, such as the Shiga toxin (*stx*) of Shigella, are primarily transferable through lateral gene transfer, which leads to the continuous evolution of pathogen species [20]. Therefore, it is necessary to examine the pathogen diversity in environmental metagenomes using essential virulence genes as biomarkers.

In this study, a toxin-centered virulence factors database was built, and the well-developed Local BLASTP method was applied to detect virulence factors in various environments. This procedure is metagenome-based and can be conducted in a high-throughput fashion, which greatly simplifies development of precautions for pathogen-enriched environments.

Methods and materials

Environments and their metagenomes

Sixty-nine metagenomes were selected and downloaded from the MG-RAST server (Table 1). These metagenomes were derived from ocean water, freshwater, wastewater, natural soil, deserts, and feces, representing the major environmental media found worldwide. Sequencing methods of the metagenomes include the illumina, Ion Torrent and 454 platforms, and predicted proteins in the metagenomes ranged from 33,743 (fresh water, ID mgm4720261) to 11,587,259 (grassland soil, ID mgm 4623645). The gene-calling results from the MG-RAST server were used for toxin factor screening in this study. The taxonomic composition at the genus level was also retrieved from the MG-RAST server for 27 representative metagenomes.

Toxin factor database

A toxin-centered database was established for bacterial pathogen detection in metagenomes in this study. Candidate toxin factors for pathogenic screening of environmental metagenomes were gathered based on well-studied pathogens summarized in the Virulence Factor Database [21], a soil borne pathogen report by Jeffery and van der Putten [2], and a manure pathogen report by the United States Water Environment Federation [22]. Sequences of the toxin factors were then retrieved by searching the UniProt database using the toxin plus pathogen names as an entry [23], while typical homologs at a cut-off E value of 10^{-6} were gathered from GenBank based on BLAST results. A protein database was then built for Local BLSATP study (Additional file 1). Considering that virulence process involves several essential factors including toxins, various pathogen-derived secretion proteins were also included in the database, and it was tested that whether secretion proteins were as specific as toxin proteins for pathogen detection. The disease relevance of all virulence factors was screened using the WikiGenes system [24] and relevant publications (Table 2).

Local BLASTP

The Local BLASTP was applied following the procedure used in our previous study [58, 59]. Basically, the gene-calling results of each metagenome were searched against the toxin factor database using BLASTP. The cut-off expectation E value was set as 10^{-6}. The results of the Local BLASTP were then copied to an Excel worksheet, after which they were subjected to duplicate removal, quality control, and subtotaled according to database ID. Duplicate removal was based on the hypothesis that each sequence contains one copy of a specific toxin factor, since the gene-calling results were used. For quality control of the BLAST results, a cut-off value of 40% for identity and 20 aa [1/3 of the length of the shortest toxin factors (e.g., the Heat-Stable Enterotoxin C)] for query alignment length were used to filter the records. The toxins abundance matrix was formed for subsequent analyses.

Table 1 General information regarding the metagenomes retrieved from the MG-RAST server

MG-RAST ID	Abbreviation	bp count	Sequence count	Protein predicted	Material	Location	Country	Coordinates	Sequencing method
mgm4440281	MW281	35,439,683	334,386	227,038	Mine drainage	Soudan Mine	USA	47.819, −92.243	454
mgm4443754	WW754	24,210,189	54,344	58,182	Waste water	Singapore	Singapore	1.332, 103.754	Unknown
mgm4443755	WW755	30,035,399	57,827	62,600	Waste water	Singapore	Singapore	1.332, 103.755	Unknown
mgm4443756	WW756	29,292,896	61,672	65,312	Waste water	Singapore	Singapore	1.332, 103.756	Unknown
mgm4461675	FW675	156,809,137	389,864	258,015	Well water	Lashkardi village	Bangladesh	23.744, 90.606	454
mgm4507016	F016	163,648,718	227,551	250,347	Feces	Bologna	Italy	44.495, 11.343	Illumina
mgm4510939	FW939	76,571,419	97,768	106,497	Surface water	Minnesota	USA	43.000, −91.000	Illumina
mgm4510941	FW941	49,162,469	56,117	68,954	Surface water	Minnesota	USA	47.000, −95.000	Illumina
mgm4510942	FW942	26,796,913	42,585	49,581	Surface water	Minnesota	USA	45.000, −94.000	Illumina
mgm4510943	FW943	52,372,093	74,178	86,925	Surface water	Minnesota	USA	45.000, −94.000	Illumina
mgm4510944	FW944	53,884,243	72,536	84,028	Surface water	Minnesota	USA	44.000, −93.000	Illumina
mgm4510945	FW945	54,142,157	71,478	83,751	Surface water	Minnesota	USA	44.000, −93.000	Illumina
mgm4510946	FW946	49,216,363	74,745	83,253	Surface water	Minnesota	USA	43.000, −93.000	Illumina
mgm4514299	DS299	322,114,449	242,0832	1,323,378	Saline Desert Soil	Gujarat of India	India	23.7925, 71.008	Ion torrent
mgm4533707	OW707	134,833,790	668,257	508,217	Brackish Water	Columbia River coastal margin, OR & WA	USA	46.265, −123.999	Illumina
mgm4543019	DS019	282,578,916	2,016,127	842,475	Saline Desert Soil	Gujarat of India	India	23.908, 70.538	Ion torrent
mgm4546371	WW371	84,424,005	907,785	803,682	Wastewater	Universiti Teknologi Malaysia	Malaysia	2.558, 104.642	Illumina
mgm4556493	AS493	162,926,938	1,613,138	1,332,131	Activated sludge	Unkown	China	31.550, 120.315	Illumina
mgm4556497	AS497	162,926,938	1,613,138	1,278,907	Activated sludge	Unkown	China	31.551, 120.316	Illumina
mgm4556505	AS505	257,249,323	2,547,023	1,737,663	Activated sludge	Unkown	China	31.552, 120.316	Illumina
mgm4556509	AS509	41,926,615	415,115	330,475	Activated sludge	Unkown	China	31.552, 120.317	Illumina
mgm4560423	F423	22,734,940	73,479	76,569	Feces	Lake Eyasi, Tanzania	Tanzania	−3.635, 35.083	Illumina
mgm4568577	MW577	10,065,266	50,137	34,287	Mine water	Guangdong	China	24.503, 113.710	454
mgm4568580	MW580	12,911,442	62,018	36,461	Mine water	Guangdong	China	22.940, 112.050	454
mgm4589537	FW537	337,068,782	2,099,471	1,842,975	Surface water	West Virginia	USA	38.094, −81.959	Illumina
mgm4620487	WW487	147,523,219	696,132	640,283	Wastewater	Guelph ON Canada	Canada	43.544, −80.248	Illumina
mgm4620488	WW488	115,131,556	578,337	537,267	Wastewater	Guelph ON Canada	Canada	43.545, −80.248	Illumina

Table 1 (continued)

MG-RAST ID	Abbreviation	bp count	Sequence count	Protein predicted	Material	Location	Country	Coordinates	Sequencing method
mgm4620491	BS491	52,759,415	244,855	238,630	Biosolides	Guelph ON Canada	Canada	43.545, −80.248	Illumina
mgm4623645	S645	1,541,232,730	15,259,730	11,587,259	Grassland soil	MPG_Ranch	USA	46.682, −114.027	Illumina
mgm4626292	S292	711,577,500	7,115,775	5,992,554	Mountain soil	Taishan, Guangdong, China	China	22.110, 112.770	Illumina
mgm4629146	D146	1,693,361,056	9,120,345	3,914,215	Dust	Valencia	Spain	39.466, −0.366	Illumina
mgm4654022	S022	830,740,317	4,484,452	4,110,948	Mountain soil	northern Galilee Mountains	Israel	33.000, 35.233	Illumina
mgm4654023	S023	711,633,780	3,762,792	3,377,414	Soil	Terra Rossa	Israel	31.700, 35.050	Illumina
mgm4654025	S025	698,042,789	3,808,872	3,379,500	Soil	Terra Rossa	Israel	31.700, 35.050	Illumina
mgm4654028	S028	572,066,482	3,129,422	2,798,806	Soil	Rendzina	Israel	31.033, 34.9	Illumina
mgm4679248	S248	603,919,746	3,365,512	1,361,948	Soil	Seoul	South Korea	37.460, 126.948	Illumina
mgm4679254	S254	689,019,062	3,688,750	1,966,121	Soil	Seoul	South Korea	37.459, 126.948	Illumina
mgm4695622	PFW622	114,430,648	111,889	148,833	Polluted fresh water	Nanjing, Jiangsu	China	32.600, 118.160	Illumina
mgm4695626	PFW626	86,732,360	78,621	111,489	Polluted fresh water	Nanjing, Jiangsu	China	32.400, 118.140	Illumina
mgm4697397	OS397	143,214,978	397,067	299,940	Organic Soil	Beijing	China	32.054, 118.763	Illumina
mgm4713197	OW197	60,417,678	272,918	140,195	Ocean water	Moorea	Pacific Ocean	17.538, −149.829	Illumina
mgm4713202	OW202	89,726,117	442,552	254,139	Ocean water	Moorea	Pacific Ocean	17.538, −149.829	Illumina
mgm4713205	OW205	106,474,596	476,363	235,777	Ocean water	Moorea	Pacific Ocean	17.538, −149.829	Illumina
mgm4718752	F752	329,518,322	1,312,822	950,489	Feces	Upstate NY	USA	42.668, −76.528	Illumina
mgm4719940	OW940	360,335,259	1,425,556	1,023,445	Ocean water	Irish Sea	Atlantic Ocean	53.225, −4.159	Illumina
mgm4720261	FW261	35,487,527	6,896	33,743	Fresh Water	Galway	Ireland	53.276, −9.060	Illumina
mgm4740560	MT560	739,577,348	7,322,548	6,102,709	Mine tailing	Kamloops	Canada	Unkown	Illumina
mgm4763187	F187	1,605,883,158	10,011,241	3,022,645	Feces	Palmerston North	New Zealand	−40.355, 175.612	Illumina
mgm4763293	F293	1,621,471,323	9,844,890	3,391,304	Feces	Palmerston North	New Zealand	−40.355, 175.612	Illumina
mgm4763371	F371	1,455,357,314	9,241,610	2,388,304	Feces	Palmerston North	New Zealand	−40.355, 175.612	Illumina
mgm4770614	S614	55,225,990	132,807	144,038	Forest soil	Leningrad	Russia	59.549, 31.399	Illumina
mgm4779571	DS571	226,100,954	1,707,504	1,288,831	Desert soil	Kutch Desert	India	23.941, 70.188	Ion Torrent
mgm4779573	OW573	126,222,705	403,606	341,516	Saline water	Lake Tyrrel	Australia	−35.32, 142.8	454
mgm4779575	OW575	46,134,631	135,959	130,566	Saline water	Albufera	Spain	39.332, −0.352	454
mgm4779577	OW577	26,544,332	80,951	44,824	Saline water	Lake Tyrrel	Australia	−35.320, 142.800	454

Table 1 (continued)

MG-RAST ID	Abbreviation	bp count	Sequence count	Protein predicted	Material	Location	Country	Coordinates	Sequencing method
mgm4779580	OW580	10,257,469	80,645	60,571	Saline water	British_ Columbia	Canada	49.730, −119.874	Ion Torrent
mgm4779582	OW582	236,630,443	691,427	455,629	Saline water	Santa Pola	Spain	38.200, −0.600	454
mgm4779585	SE585	69,798,594	852,769	442,408	Sediment	Yilgarn_Craton	Australia	−33.426, 121.689	Illumina
mgm4779589	OW589	10,713,143	81,307	61,739	Saline water	British Columbia	Canada	49.730, −119.874	Ion torrent
mgm4784118	SE118	294,673,007	263,221	400,006	Sediment	St. Francis Bay	South Africa	−34.190, 24.704	Ion Torrent
mgm4784267	SE267	337,590,087	280,494	451,716	Sediment	Cape Recife	South Africa	−34.045, 25.569	Ion Torrent
mgm4795328	FW328	1,263,293,418	5,301,996	2,970,195	Surface water	San Antonio	USA	29.424, −98.493	Illumina
mgm4819059	S059	658,065,701	4,358,051	2,432,031	Forest soil	Lahti	Finland	60.971, 25.704	Illumina
mgm4819062	S062	1,197,697,874	7,931,774	6,478,848	Forest soil	Baltimore	USA	39.488, −76.689	Illumina
mgm4819067	S067	771,429,857	5,108,807	3,703,810	Grassland soil	Potchefstroom	South Africa	−26.701, 27.101	Illumina
mgm4819073	S073	730,894,511	4,840,361	3,946,449	Forest soil	Baltimore	USA	39.326, −76.622	Illumina
N/A	SRS	532,850,584	1,632,914	1,408,943	Red Soil	Mt Isa	Australia	20.440, 139.300	Illumina
N/A	SRSP	433,386,397	1,338,665	1,081,822	Red soil polluted	Mt Isa	Australia	20.440, 139.300	Illumina
N/A	SSLS	507,124,889	1,552,234	1,413,889	Shrub land soil	Mt Isa	Australia	20.440, 139.300	Illumina

Specificity tests of the Local BLASTP method

Sequences from the toxin database established in this study, as "known sequences" to the database, were selected randomly and searched against the database using the BLASTP procedure. The genome of *Clostridium perfringens* ATCC 13124 (NC_008261), as "unknown" sequences to the database, was subject to the Local BLASTX procedure as well. Homologous proteins were searched exhaustively in the GenBank database using BLASTP, with the representative toxin factors in the toxins database as a query. Sequences were retrieved and aligned using ClustalW [60], and Maximum-likelihood phylogeny was conducted with MEGA 7 [61].

Data analysis

The toxin frequency in each metagenome was normalized to a total gene frequency of 10,000,000 to eliminate the effects of gene pool size. Toxin abundance in the 69 metagenomes was visualized using Circos [62]. The genus abundance of 27 selected metagenomes representing the main environment types was calculated and sorted by genus name, followed by manual construction of a genus abundance matrix for subsequent biodiversity-toxin abundance Canonical Correspondence Analysis using R [63] with the package 'vegan' [64].

Results and discussion

In this study, a toxin-centered database was built for bacterial pathogen screening in various microbiomes through a Local BLASTP procedure. The specificity of the procedure was tested, the relative abundance of toxins in the microbiomes was examined, and the toxin-taxonomic abundance correspondence analysis was performed.

Like the previously established Local BLASTN method for antibiotic and metal resistance genes screening [58, 59, 65], the Local BLASTP method using the toxin-centered pathogen database in this study was successful at accurately identifying toxin proteins from the database. For screening of the *Clostridium perfringens* ATCC 13124 genome, the methods successfully detected the pore-forming genes and multiple copies of the glucosyltransferase (*toxB*-like) and ADP-ribosyltransferase (*spvB*-like) genes, based on the raw data. These results are consistent with the virulence genetic features of *Clostridium* sp. [21], which have not been well detailed

Table 2 Typical virulence factors investigated in this study and their disease–relevance

Toxin factor	ID in the databse	Typical pathogens and (disease)	Role of the toxin	Reference
Aerolysin	aerA	*Aeromonas* spp.	Cytolytic pore-forming	[25]
Alveolysin	alo	*Bacillus anthracis*	Pore-forming	[26]
Dermonecrotic toxin	dnt	*Bordetella pertussis*	Stimulating the assembly of actin stress fibers and focal adhesions	[27, 28]
Pertussis toxin subunit 1	ptxA	*Bo. pertussis*	Causing disruption of host cellular regulation	[28]
Type IV secretion system protein Ptl	ptlCH	*Bo. pertussis*	Secretion of pertussis toxin	[28]
Chlamydia protein associating with death domains	CADD	*Chlamydia trachomatis* (trachoma, urethritis, etc.)	Inducing cell apoptosis	[29]
Perfringolysin O	pfo	*Clostridium perfringens* (food poisoning)	Pore-forming	[30]
Glucosyltransferase toxin B	toxB	*Cl. sordellii* *Cl. difficile* (diarrhea)	Cytopathic effects	[31]
Zeta toxin family protein	ZETA	*Coxiella* sp. DG_40	Inhibiting cell wall biosynthesis	[32]
Shiga toxin 1	stx1	*Escherichia coli* (diarrhea) *Shigella dysenteriae* (Shigellosis)	Haemolytic uraemic syndrome	[20]
Toxin CdiA	cdiA	*E. coli* *Yersinia pestis* (plaque)	Decreasing aerobic respiration and ATP levels	[33]
Shiga-like toxin 2	stx2	*Enterobacteria phage* 933 W *E. coli*	Haemolytic uraemic syndrome	[34]
Repeats-in toxin	rtxA	*Legionella pneumophila* (Legionnaries' disease) *Aeromonas dhakensis* (gastroenteritis, septicemia)	Adherence and pore forming	[35, 36]
Cholera toxin secretion protein EpsF	epsF	*Le. pneumophila*	Toxin secretion	[37]
Toxin secretion ATP binding protein	LwT1SS	*Le. waltersii*	Toxin secretion	[38]
1-phosphatidylinositol phosphodiesterase	PLC	*Listeria monocytogenes* (listeriosis)	Lysis of the phagolysosomal membrane	[39]
Listeriolysin O	hly	*Li. monocytogenes* (listeriosis)	Pore forming, hemolysin	[40, 41]
Outer membrane channel protein CpnT	cpnT	*Mycobacterium tuberculosis*	Nutrient uptake	[42]
RNA interferase	mazF/pemK/ndoA/relE/relK/relG/yoeB/higB/mvpA	*Proteus vulgaris* (wound infections) *My. tuberculosis* (tuberculosis) *E. coli*	Cleavage of cellular mRNAs, inhibiting growth	[43–48]
Hemolytic phospholipase C	plcH	*Pseudomonas Aeruginosa* *Cl. perfringens* (food poisoning)	Membrane-damaging	[49]
ADP-ribosyltransferase toxin	exoS	*Ps. aeruginosa*	Inhibition of phagocytosis	[50]
Ornithine carbamoyltransferase	argK	*Ps. savastanoi*	Promoting survival and pathogenicity	[51]
Exoenzyme U	exoU	*Ps. aeruginosa*	Membrane-lytic and cytotoxic	[52]
Exotoxin A	ETA	*Ps. aeruginosa* (eye and wound infections)	ADP-ribosylating eukaryotic elongation factor 2	[53]
Mono(ADP-ribosyl)transferase	spvB	*Salmonella dublin* (gastroenteritis)	ADP-ribosylating, destabilizing cytoskeleton	[54]
Adenylate cyclase	cyaA	*Sa. choleraesuis* (typhoid fever) *Bo. pertussis* (whooping cough) *Ba. anthracis* (anthrax)	Pore-forming with cAMP-elevating activity	[55]
Endonuclease VapC	vapC	*Shigella flexneri* (diarrhea) *Sa. Dublin* *My. Tuberculosis* *Coxiella burnetii* (Q fever)	tRNase activities	[45]
Leucotoxin	luk	*Staphylococcus aureus* (sinusitis, skin abscess)	Lysis of leukocytes	[56]
Exfoliative toxin	ET	*St. aureus*	Proteolytic activity	[57]

in the GenBank annotation record. Such a cross-validation positively indicated that the Local BLASTP procedure established here is useful in predicting toxin genes in unknown genomes. Yet for a semi-quantitative method to estimate toxin factors in metagenomes, a false positive analysis is required to examine to what level mismatch is included in the Local BLASTP results. Actually, the cut-off values of identity greatly impact the homolog virulence factor abundance returned. At cut-off values of 40% for identity and 20 aa for alignment length, only four records for *Clostridium perfringens* ATCC 13124 genome query were returned after duplication removal, one for 1-phosphatidylinositol phosphodiesterase, one for pore-forming alveolysin, one for Ornithine carbamoyltransferase and one for RNA interferase NdoA. At a cut-off identity value of 35%, one more record (Toxin secretion ATP binding protein) was returned. This means that the Local BLASTP procedure was able to detect the virulence factors in unknown genomic dataset at least semi-quantitatively, with proper cut-off values for data quality control. The accuracy of the BLASTP procedure in virulence factor detection was further tested using the genomes of *Bacillus thuringiensis serovar konkukian* str. 97-27 (AE017355.1) and *Helicobacter pylori* 26695 (AE000511.1).

As mentioned above, functional genes including toxin factors may partly evolve through lateral gene transfer, which makes their taxonomic affiliation difficult. It is thus interesting to explore how specific toxin factors are associated with the taxonomic units of pathogens. Here, I explored this issue by investigating the taxonomic distribution of homologs of toxins retrieved from the GenBank database. Generally, at a lower expectation value, most toxins were associated with a specific group of pathogens. For example, at the default cut-off E value, 241 out of 242 returned records of *Mycobacterium tuberculosis* RelE homologs fell within the phylum *Actinobacteria*. Moreover, 89% of these homologs were from the genus *Mycobacterium*, while 99.7% of *Yersinia pestis* CdiA homologs and 92.7% of *Bordetella pertussis* cya homologs belonged to *Proteobacteria*, and homologs of *Aeromonas dhakensis* repeats-in toxin (RtxA) were mostly associated with the class *Gammaproteobacteria* (206 out of 242). However, no obvious genus-toxin association was identified. It is worth noting that these results largely depended on the availability of toxin sequences in each taxonomic unit. The lack of a genus-toxin association basically denied the possibility of detecting a specific pathogen using a specific toxin as a single signature.

It is still not clear whether virulence secretion proteins are specific for pathogen detection as signatures, though they are essential for virulence process [20]. For example, the contact-dependent toxin delivery protein CdiA

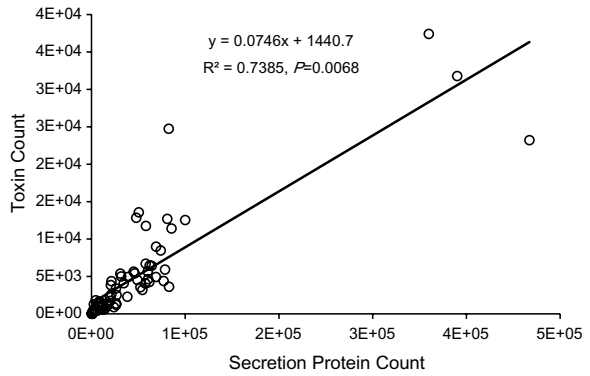

Fig. 1 Correlation between relative abundance of toxins and secretion proteins in the global microbiomes (N = 69)

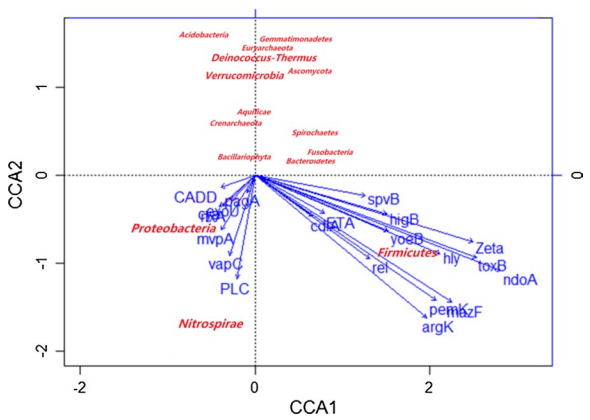

Fig. 2 Canonical correspondence analysis of the associations between phyla and toxins from typical environments

was found to be widespread in bacteria [33]. The relative abundance of secretion proteins in the 69 microbiomes was determined as well as that of the toxins which are essential to virulence processes. The results of the present study showed that the abundance of secretion proteins selected in the database was strongly correlated with the toxin abundance ($R^2 = 0.74$, $P = 0.0068$, Fig. 1). The most abundant secretion proteins included *L. waltersii* toxin secretion protein (LWT1SS), *L. pneumophila* toxin secretion protein ApxIB, and *Aeromonas hydrophila* RTX transporter (RtxB) (data not shown). Further exploration indicated that although *A. hydrophila* RtxB homologs from GenBank were found in all *Proteobacteria* classes, most of the RtxB-harboring species have been reported to be pathogens, including *Vibrio* spp. [64, 66], *Pseudomonas* spp., *Neisseria meningitides* [67], *Ralstonia* spp. [68], and *Yersinia* spp. [21]. This may imply the pathogen-specific nature of secretion proteins included in the database, and that toxin secretion proteins can be used as signatures for pathogen detection as well.

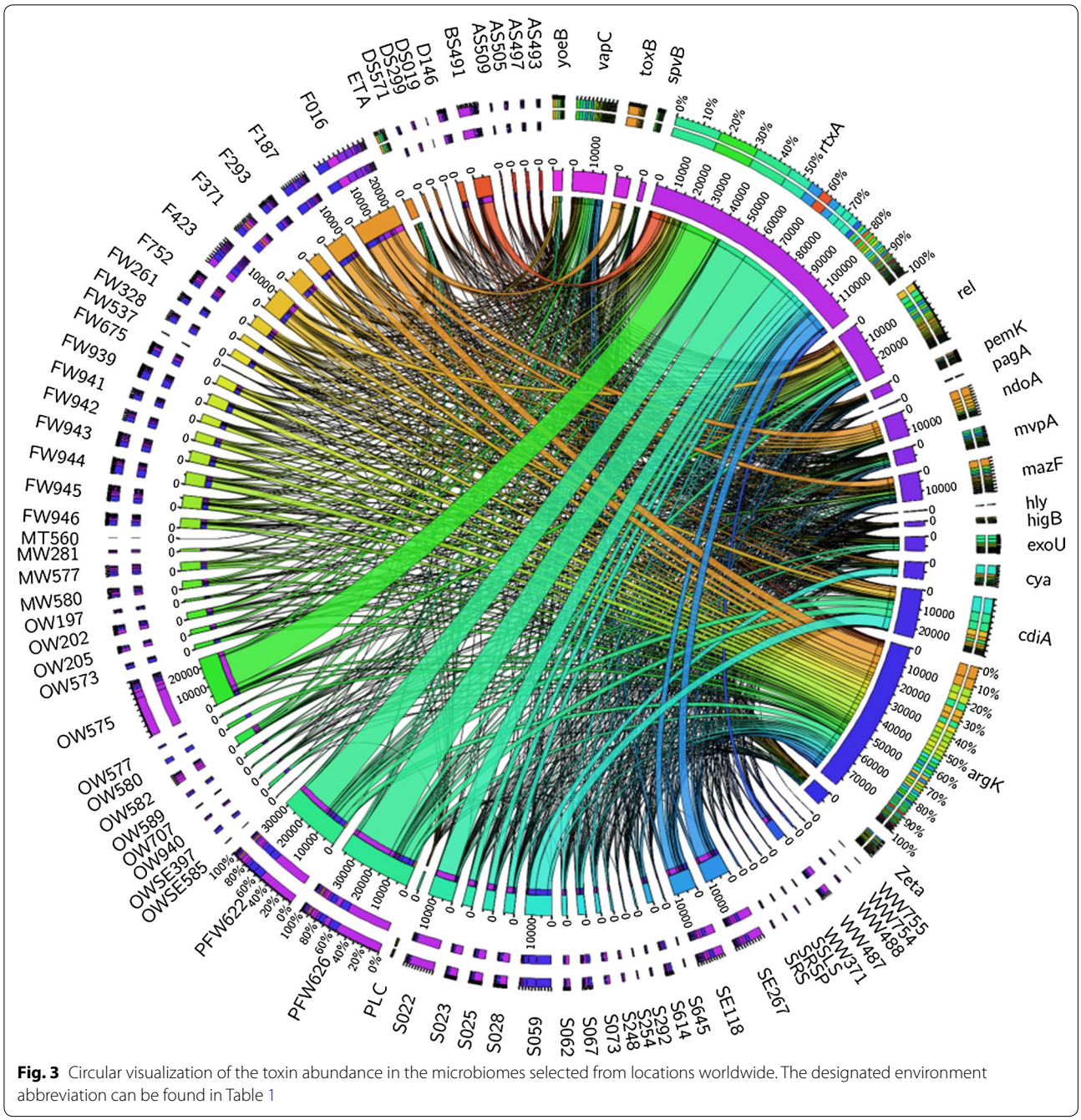

Fig. 3 Circular visualization of the toxin abundance in the microbiomes selected from locations worldwide. The designated environment abbreviation can be found in Table 1

Toxin-phyla CCA results showed that all phyla can be clearly separated into two groups, and that almost all toxins were associated with *Proteobacteria*, *Nitrospirae,* and *Firmicutes* (Fig. 2). Considering the phylum-specificity of the toxins stated above, these results can be biased because of the taxonomic affiliation of toxins included in the Local BLASTP database. The taxonomic distribution proportion of currently available genomes of identified pathogens was reflected in the toxin database, with *Proteobacteria* and *Firmicutes* accounting for the majority of the genomes. However, the CCA results may also indicate, at least in part, a proportional lack of pathogens in some phyla, such as *Crenarchaeota*, *Euryarchaeota*, *Verrucomicrobia,* and *Bacteroidetes* [69].

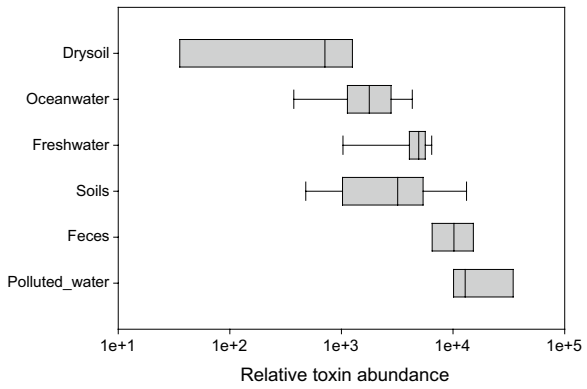

Fig. 4 A Boxplot showing the relative abundance of toxins detected from the metagenomes in this study. Drysoil includes the desert soil and desert mine tailings

Archaea cannot easily absorb phage particles because of their extracellular structures, which differ from bacteria [70]. A recent study by Li et al. [9] also found that the five most abundant bacterial pathogens were from either *Proteobacteria* or *Firmicutes* in wastewater microbiomes. Taken together, these findings could indicate that *Proteobacteria* or *Firmicutes* were evolutionarily enriched with pathogens when they dominated most environmental microbiomes on the planet [71, 72].

Interestingly, there was a strong association between the phylum *Nitrospirae* and toxins of RNase inteferases (MvpA and MapC) and *Listeria monocytogenes*1-phosphatidylinositol phosphodiesterase PLC. Further searches against the UniProt database [73] revealed no homologous records of MvpA and PLC from *Nitrospirae*, and only 109 out of 15,574 bacterial records for VapC were from *Nitrospirae*. These findings imply that there may be many more *Nitrospirae* pathogens harboring MvpA and PLC that have yet to be discovered.

The screening of toxins in the 69 global microbiomes revealed the most prevalent toxins and pathogen-enriched environments. Specifically, the results showed that pore-forming RTX toxin and ornithine carbamoyltransferase ArgK were most prevalent globally in terms of both occurrence and relative abundance (Fig. 3). RTX toxins comprise a large family of pore-forming exotoxins. Known homologs in the GenBank database of *Aeromonas dhakensis* RtxA were mainly in the genera of *Aeromonas*, *Pseudomonas* (e.g., CP015992), *Vibrio* (e.g., CP002556), and *Legionella* (e.g., CP015953). These genera are well known to be associated with gastroenteritis, eye and wound infections, cholera and legionellosis, and RTX toxins are a key part of the virulence systems of each of these conditions [74–77]. The *argK* gene is a part of the Pht cluster, which contains genes for the synthesis of phaseolotoxin in *Ps. syringae* pv. phaseolicola [78]. ArgK plays an essential role in the survival and pathogenicity of *Ps. syringae*. Known ArgK proteins mainly come from *Pseudomonas*, *Escherichia,* and *Mycobacterium*, which are widespread and persistent in the environment [79]. In addition, Cya is worth noting as an essential unit of *Bacillus anthracis* virulence that causes anthrax and may lead to mammalian death [80]. Known homologs in the GenBank database of *Bacillus anthracis* Cya were mainly from *Bacillus* spp., *Bordetella* spp., *Pseudomonas aeruginosa*, *Yersinia pseudotuberculosis*, and *Vibrio* spp. Their presence in the environment should be carefully examined and precautions should be taken to prevent infection by these organisms since many of them are associated with very common diseases such as whooping cough.

The main purpose of the Local BLASTP method established here was to screen pathogen-enriched environments to enable development of precautionary measures. Our results clearly indicated that contaminated freshwater, feces, and harbor sediment microbiomes were rich in pathogens (Fig. 4). Although there was no detailed background information regarding these environments in this study, the results presented herein may provide important implications for pathogen-related risk control. Surprisingly, two lake water microbiomes from Nanjing, China contained the highest toxin factors among the 69 samples. Further investigation of the location and contamination status supported the sewage-nature of the lake water. In China, most polluted lakes receive sewage that includes feces materials [81]. According to an official survey conducted in 2015, Nanjing has 28 lakes with a total area of 14 km^2, among which 96.4% are classified as polluted (Class V of the national standard). Studies have documented that pathogens tend to be enriched in polluted waters [13]. It is not surprising to find that feces samples had very high abundance of toxins. Epidemical statistics have indicated that feces are the most important pathway for diarrheal diseases, which is a leading cause of childhood death globally [82]. Meanwhile, dry soil environments like desert soil and desert tailings were found to contain relatively less toxin factors. It is still unclear to what extent the environments stressed by long-lasting drought or metal pollution suppress the colonization and development of pathogens [83]. In all, the association between environmental factors and pathogen abundance merits a systematic exploration in the future.

Conclusions

A Local BLASTP procedure was established for rapid detection of toxins in environmental samples. Screening of global microbiomes in this study provided a

quantitative estimate of the most prevalent toxins and most pathogen-enriched environments.

Additional file

Additional file 1. A toxin factor database for metagenomic detection of environmental pathogens through Local BLASTP.

Abbreviations

BLAST: basic local alignment search tool; PCR: polymerase chain reaction.

Acknowledgements

I thank Dr. Philip L. Bond and The University of Queensland for providing training in bioinformatics. I would like to thank LetPub (http://www.letpub.com) for providing linguistic assistance during the preparation of this manuscript. I also thank the founders of the existing pathogen-relevant database, particularly the Virulence Factor Database, which provided valuable reference for the build-up of the toxin database in this study.

Authors' contributions

XL initiated the study, analyzed the data, and wrote the manuscript. The author read and approved the final manuscript.

Funding

This work was financially supported by the National Key Research and Development Program of China (2018YFD0800306), the National Natural Science Foundation of China (41877414), and Hebei Science Fund for Distinguished Young Scholars (D2018503005).

Availability of data and materials

The toxin database is available in the Additional file 1: Materials. All toxin abundance data in this study can be provided by the author upon request.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Received: 22 February 2019 Accepted: 22 May 2019
Published online: 11 June 2019

References

1. Baumgardner DJ (2012) Soil-related bacterial and fungal infections. J Am Board Fam Med 25(5):734–744. https://doi.org/10.3122/jabfm.2012.05.110226
2. Jeffery S, van der Putten WH (2011) Soil borne diseases of Humans. Joint Research Centre, Institute for Environment and Sustainability Ispra, Italy
3. Whiley H, Bentham R (2011) Legionella longbeachae and Legionellosis. Emerg Infect Dis 17(4):579–583. https://doi.org/10.3201/eid1704.100446
4. Cooper RC, Golueke CG (1979) Survival of enteric bacteria and viruses in compost and its leachate. Compost Sci Land Ut 20(2):29–35
5. Dan TBB, Wynne D, Manor Y (1997) Survival of enteric bacteria and viruses in Lake Kinneret, Israel. Water Res 31(11):2755–2760
6. Keswick BH, Gerba CP, Secor SL, Cech I (1982) Survival of enteric viruses and indicator bacteria in groundwater. J Environ Sci Health A 17(6):903–912
7. Wait DA, Sobsey MD (2001) Comparative survival of enteric viruses and bacteria in Atlantic Ocean seawater. Water Sci Technol 43(12):139–142
8. Miller RR, Montoya V, Gardy JL, Patrick DM, Tang P (2013) Metagenomics for pathogen detection in public health. Genome Med 5:81. https://doi.org/10.1186/gm485
9. Li B, Ju F, Cai L, Zhang T (2015) Profile and fate of bacterial pathogens in sewage treatment plants revealed by high-throughput metagenomic approach. Environ Sci Technol 49(17):10492–10502. https://doi.org/10.1021/acs.est.5b02345
10. Amha YM, Anwar MZ, Kumaraswamy R, Henschel A, Ahmad F (2017) Mycobacteria in municipal wastewater treatment and reuse: microbial diversity for screening the occurrence of clinically and environmentally relevant species in arid regions. Environ Sci Technol 51(5):3048–3056. https://doi.org/10.1021/acs.est.6b05580
11. Granberg F, Vicente-Rubiano M, Rubio-Guerri C, Karlsson OE, Kukielka D, Belak S, Sanchez-Vizcaino JM (2013) Metagenomic detection of viral pathogens in Spanish honeybees: co-infection by aphid lethal paralysis, Israel acute paralysis and Lake Sinai viruses. PLoS ONE 8(2):e57459. https://doi.org/10.1371/journal.pone.0057459
12. Yang J, Yang F, Ren LL, Xiong ZH, Wu ZQ, Dong J, Sun LL, Zhang T, Hu YF, Du J, Wang JW, Jin Q (2011) Unbiased parallel detection of viral pathogens in clinical samples by use of a metagenomic approach. J Clin Microbiol 49(10):3463–3469. https://doi.org/10.1128/Jcm.00273-11
13. Bibby K (2013) Metagenomic identification of viral pathogens. Trends Biotechnol 31(5):11–15. https://doi.org/10.1016/j.tibtech.2013.01.016
14. Nakamura S, Yang CS, Sakon N, Ueda M, Tougan T, Yamashita A, Goto N, Takahashi K, Yasunaga T, Ikuta K, Mizutani T, Okamoto Y, Tagami M, Morita R, Maeda N, Kawai J, Hayashizaki Y, Nagai Y, Horii T, Iida T, Nakaya T (2009) Direct metagenomic detection of viral pathogens in nasal and fecal specimens using an unbiased high-throughput sequencing approach. PLoS ONE 4(1):e4219. https://doi.org/10.1371/journal.pone.0004219
15. Fukui Y, Aoki K, Okuma S, Sato T, Ishii Y, Tateda K (2015) Metagenomic analysis for detecting pathogens in culture-negative infective endocarditis. J Infect Chemother 21(12):882–884. https://doi.org/10.1016/j.jiac.2015.08.007
16. Baldwin DA, Feldman M, Alwine JC, Robertson ES (2014) Metagenomic assay for identification of microbial pathogens in tumor tissues. Mbio 5(5):e01714–14. https://doi.org/10.1128/mbio.01714-14
17. Achtman M, Wagner M (2008) Microbial diversity and the genetic nature of microbial species. Nat Rev Microbiol 6(6):431–440
18. Liao PY, Lee KH (2010) From SNPs to functional polymorphism: the insight into biotechnology applications. Biochem Eng J 49(2):149–158. https://doi.org/10.1016/j.bej.2009.12.021
19. Sokurenko EV, Hasty DL, Dykhuizen DE (1999) Pathoadaptive mutations: gene loss and variation in bacterial pathogens. Trends Microbiol 7(5):191–195. https://doi.org/10.1016/S0966-842x(99)01493-6
20. Strauch E, Lurz R, Beutin L (2001) Characterization of a Shiga toxin-encoding temperate bacteriophage of Shigella sonnei. Infect Immun 69(12):7588–7595. https://doi.org/10.1128/Iai.69.12.7588-7595.2001
21. Chen LH, Zheng DD, Liu B, Yang J, Jin Q (2016) VFDB 2016: hierarchical and refined dataset for big data analysis-10 years on. Nucleic Acids Res 44(D1):D694–D697. https://doi.org/10.1093/nar/gkv1239
22. Water Environment Federation (2009) Manure pathogens: manure management, regulations, and water quality protection. WEF Press, Water Environment Federation, Alexandria
23. Bateman A, Martin MJ, O'Donovan C, Magrane M, Alpi E, Antunes R, Bely B, Bingley M, Bonilla C, Britto R, Bursteinas B, Bye-A-Jee H, Cowley A, Da Silva A, De Giorgi M, Dogan T, Fazzini F, Castro LG, Figueira L, Garmiri P, Georghiou G, Gonzalez D, Hatton-Ellis E, Li WZ, Liu WD, Lopez R, Luo J, Lussi Y, MacDougall A, Nightingale A, Palka B, Pichler K, Poggioli D, Pundir S, Pureza L, Qi GY, Rosanoff S, Saidi R, Sawford T, Shypitsyna A, Speretta E, Turner E, Tyagi N, Volynkin V, Wardell T, Warner K, Watkins X, Zaru R, Zellner H, Xenarios I, Bougueleret L, Bridge A, Poux S, Redaschi N, Aimo L, Argoud-Puy G, Auchincloss A, Axelsen K, Bansal P, Baratin D, Blatter MC, Boeckmann B, Bolleman J, Boutet E, Breuza L, Casal-Casas C, de Castro E, Coudert E, Cuche B, Doche M, Dornevil D, Duvaud S, Estreicher A, Famiglietti L, Feuermann M, Gasteiger E, Gehant S, Gerritsen V, Gos A, Gruaz-Gumowski N, Hinz U, Hulo C, Jungo F, Keller G, Lara V, Lemercier P, Lieberherr D, Lombardot T, Martin X, Masson P, Morgat A, Neto T, Nouspikel N, Paesano S, Pedruzzi I, Pilbout S, Pozzato M, Pruess M, Rivoire C, Roechert B, Schneider M, Sigrist C, Sonesson K, Staehli S, Stutz A, Sundaram S, Tognolli M, Verbregue L, Veuthey AL, Wu CH, Arighi CN, Arminski L, Chen CM, Chen YX, Garavelli JS, Huang HZ, Laiho K, McGarvey P, Natale

DA, Ross K, Vinayaka CR, Wang QH, Wang YQ, Yeh LS, Zhang J, Consortium U (2017) UniProt: the universal protein knowledgebase. Nucleic Acids Res 45(D1):D158–D169. https://doi.org/10.1093/nar/gkw1099
24. Hoffmann R (2008) A wiki for the life sciences where authorship matters. Nat Genet 40(9):1047–1051. https://doi.org/10.1038/ng.f.217
25. Howard SP, Garland WJ, Green MJ, Buckley JT (1987) Nucleotide sequence of the gene for the hole-forming toxin aerolysin of *Aeromonas hydrophila*. J Bacteriol 169(6):2869–2871
26. Geoffroy C, Mengaud J, Alouf JE, Cossart P (1990) Alveolysin, the thiol-activated toxin of *Bacillus alvei*, is homologous to listeriolysin O, perfringolysin O, pneumolysin, and streptolysin O and contains a single cysteine. J Bacteriol 172(12):7301–7305. https://doi.org/10.1128/jb.172.12.7301-7305.1990
27. Matsuzawa T, Kashimoto T, Katahira J, Horiguchi Y (2002) Identification of a receptor-binding domain of bordetella dermonecrotic toxin. Infect Immun 70(7):3427–3432. https://doi.org/10.1128/iai.70.7.3427-3432.2002
28. Weiss AA, Johnson FD, Burns DL (1993) Molecular characterization of an operon required for pertussis toxin secretion. Proc Natl Acad Sci U S A 90(7):2970–2974. https://doi.org/10.1073/pnas.90.7.2970
29. Schwarzenbacher R, Stenner-Liewen F, Liewen H, Robinson H, Yuan H, Bossy-Wetzel E, Reed JC, Liddington RC (2004) Structure of the chlamydia protein CADD reveals a redox enzyme that modulates host cell apoptosis. J Biol Chem 279(28):29320–29324. https://doi.org/10.1074/jbc.M401268200
30. Rossjohn J, Polekhina G, Feil SC, Morton CJ, Tweten RK, Parker MW (2007) Structures of perfringolysin O suggest a pathway for activation of cholesterol-dependent cytolysins. J Mol Biol 367(5):1227–1236. https://doi.org/10.1016/j.jmb.2007.01.042
31. Lyras D, O'Connor JR, Howarth PM, Sambol SP, Carter GP, Phumoonna T, Poon R, Adams V, Vedantam G, Johnson S, Gerding DN, Rood JI (2009) Toxin B is essential for virulence of *Clostridium difficile*. Nature 458(7242):1176–1181. https://doi.org/10.1038/nature07822
32. Lioy VS, Machon C, Tabone M, Gonzalez-Pastor JE, Daugelavicius R, Ayora S, Alonso JC (2012) The ζ toxin induces a set of protective responses and dormancy. PLoS ONE 7(1):e30282. https://doi.org/10.1371/journal.pone.0030282
33. Aoki SK, Diner EJ, de Roodenbeke CT, Burgess BR, Poole SJ, Braaten BA, Jones AM, Webb JS, Hayes CS, Cotter PA, Low DA (2010) A widespread family of polymorphic contact-dependent toxin delivery systems in bacteria. Nature 468(7322):439–442. https://doi.org/10.1038/nature09490
34. Schmidt H, Scheef J, JanetzkiMittmann C, Datz M, Karch H (1997) An ileX tRNA gene is located close to the Shiga toxin II operon in enterohemorrhagic *Escherichia coli* O157 and non-O157 strains. FEMS Microbiol Lett 149(1):39–44. https://doi.org/10.1016/S0378-1097(97)00051-7
35. D'Auria G, Jimenez N, Peris-Bondia F, Pelaz C, Latorre A, Moya A (2008) Virulence factor rtx in *Legionella pneumophila*, evidence suggesting it is a modular multifunctional protein. BMC Genomics 9:1. https://doi.org/10.1186/1471-2164-9-14
36. Rasmussen-Ivey CR, Figueras MJ, McGarey D, Liles MR (2016) Virulence factors of *Aeromonas hydrophila*: in the wake of reclassification. Front Microbiol 7:1337. https://doi.org/10.3389/fmicb.2016.01337
37. Sandkvist M, Michel LO, Hough LP, Morales VM, Bagdasarian M, Koomey M, DiRita VJ, Bagdasarian M (1997) General secretion pathway (eps) genes required for toxin secretion and outer membrane biogenesis in *Vibrio cholerae*. J Bacteriol 179(22):6994–7003. https://doi.org/10.1128/jb.179.22.6994-7003.1997
38. Söderberg MA, Rossier O, Cianciotto NP (2004) The type II protein secretion system of *Legionella pneumophila* promotes growth at low temperatures. J Bacteriol 186(12):3712–3720. https://doi.org/10.1128/jb.186.12.3712-3720.2004
39. Henner DJ, Yang M, Chen E, Hellmiss R, Rodriguez H, Low MG (1988) Sequence of the *Bacillus thuringiensis* phosphatidylinositol specific phospholipase-c. Nucleic Acids Res 16(21):10383. https://doi.org/10.1093/nar/16.21.10383
40. Cossart P (1988) The listeriolysin O-gene—a chromosomal locus crucial for the virulence of listeria-monocytogenes. Infection 16:S157–S159. https://doi.org/10.1007/Bf01639740
41. Hamon MA, Batsche E, Regnault B, Tham TN, Seveau S, Muchardt C, Cossart P (2007) Histone modifications induced by a family of bacterial toxins. Proc Natl Acad Sci U S A 104(33):13467–13472. https://doi.org/10.1073/pnas.0702729104
42. Danilchanka O, Pires D, Anes E, Niederweis M (2015) The *Mycobacterium tuberculosis* outer membrane channel protein CpnT confers susceptibility to toxic molecules. Antimicrob Agents chemother 59(4):2328–2336. https://doi.org/10.1128/Aac.04222-14
43. Hurley JM, Woychik NA (2009) Bacterial toxin HigB associates with ribosomes and mediates translation-dependent mRNA cleavage at A-rich sites. J Biol Chem 284(28):18605–18613. https://doi.org/10.1074/jbc.M109.008763
44. Korch SB, Contreras H, Clark-Curtiss JE (2009) Three *Mycobacterium tuberculosis* rel toxin-antitoxin modules inhibit mycobacterial growth and are expressed in infected human macrophages. J Bacteriol 191(5):1618–1630. https://doi.org/10.1128/jb.01318-08
45. Pullinger GD, Lax AJ (1992) A salmonella-dublin virulence plasmid locus that affects bacterial-growth under nutrient-limited conditions. Mol Microbiol 6(12):1631–1643. https://doi.org/10.1111/j.1365-2958.1992.tb00888.x
46. Tian QB, Ohnishi M, Tabuchi A, Terawaki Y (1996) A new plasmid-encoded proteic killer gene system: cloning, sequencing, and analysing hig locus of plasmid Rts1. Biochem Biophys Res Commun 220(2):280–284. https://doi.org/10.1006/bbrc.1996.0396
47. Pellegrini O, Mathy N, Gogos A, Shapiro L, Condon C (2005) The *Bacillus subtilis* ydcDE operon encodes an endoribonuclease of the MazF/PemK family and its inhibitor. Mol Microbiol 56(5):1139–1148. https://doi.org/10.1111/j.1365-2958.2005.04606.x
48. Yamaguchi Y, Inouye M (2011) Regulation of growth and death in *Escherichia coli* by toxin-antitoxin systems. Nat Rev Microbiol 9(11):779–790. https://doi.org/10.1038/nrmicro2651
49. Songer JG (1997) Bacterial phospholipases and their role in virulence. Trends Microbiol 5(4):156–161. https://doi.org/10.1016/s0966-842x(97)01005-6
50. Krueger KM, Barbieri JT (1995) The family of bacterial ADP-ribosylating exotoxins. Clin Microbiol Rev 8(1):34–47
51. Hatziloukas E, Panopoulos NJ (1992) Origin, structure, and regulation of argK, encoding the phaseolotoxin-resistant ornithine carbamoyltransferase in *Pseudomonas syringae* pv. phaseolicola, and functional expression of argK in transgenic tobacco. J Bacteriol 174(18):5895–5909. https://doi.org/10.1128/jb.174.18.5895-5909.1992
52. Phillips RM, Six DA, Dennis EA, Ghosh P (2003) In vivo phospholipase activity of the *Pseudomonas aeruginosa* cytotoxin ExoU and protection of mammalian cells with phospholipase A2 inhibitors. J Biol Chem 278(42):41326–41332. https://doi.org/10.1074/jbc.M302472200
53. Yates SP, Merrill AR (2004) Elucidation of eukaryotic elongation factor-2 contact sites within the catalytic domain of *Pseudomonas aeruginosa* exotoxin A. Biochem J 379:563–572. https://doi.org/10.1042/Bj20031731
54. Lesnick ML, Reiner NE, Fierer J, Guiney DG (2001) The Salmonella spvB virulence gene encodes an enzyme that ADP-ribosylates actin and destabilizes the cytoskeleton of eukaryotic cells. Mol Microbiol 39(6):1464–1470. https://doi.org/10.1046/j.1365-2958.2001.02360.x
55. Skopova K, Tomalova B, Kanchev I, Rossmann P, Svedova M, Adkins I, Bibova I, Tomala J, Masin J, Guiso N, Osicka R, Sedlacek R, Kovar M, Sebo P (2017) cAMP-elevating capacity of the adenylate cyclase toxin-hemolysin is sufficient for lung infection but not for full virulence of Bordetella pertussis. Infect Immun 85(6):e00937–16. https://doi.org/10.1128/iai.00937-16
56. Labandeira-Rey M, Couzon F, Boisset S, Brown EL, Bes M, Benito Y, Barbu EM, Vazquez V, Hook M, Etienne J, Vandenesch F, Bowden MG (2007) *Staphylococcus aureus* Panton-Valentine leukocidin causes necrotizing pneumonia. Science 315(5815):1130–1133. https://doi.org/10.1126/science.1137165
57. Bukowski M, Wladyka B, Dubin G (2010) Exfoliative toxins of *Staphylococcus aureus*. Toxins 2(5):1148–1165. https://doi.org/10.3390/toxins2051148
58. Li X, Zhu YG, Shaban B, Bruxner TJ, Bond PL, Huang L (2015) Assessing the genetic diversity of Cu resistance in mine tailings through high-throughput recovery of full-length copA genes. Sci Rep 5:13258. https://doi.org/10.1038/srep13258
59. Li XF, Bond PL, Huang LB (2017) Diversity of As metabolism functional genes in Pb-Zn mine tailings. Pedosphere 27(3):630–637. https://doi.org/10.1016/S1002-0160(17)60357-6
60. Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R, Thompson JD, Gibson

TJ, Higgins DG (2007) Clustal W and clustal X version 2.0. Bioinformatics 23(21):2947–2948
61. Kumar S, Stecher G, Tamura K (2016) MEGA7: Molecular evolutionary genetics analysis version 7.0 for bigger datasets. Mol Biol Evol 33(7):1870–1874. https://doi.org/10.1093/molbev/msw054
62. Krzywinski M, Schein J, Birol I, Connors J, Gascoyne R, Horsman D, Jones SJ, Marra MA (2009) Circos: an information aesthetic for comparative genomics. Genome Res 19(9):1639–1645. https://doi.org/10.1101/gr.092759.109
63. R: A language and environment for statistical computing (2016) R Foundation for Statistical Computing
64. Oksanen J, Blanchet FG, Friendly M, Kindt R, Legendre P, McGlinn D, Minchin PR, O'Hara RB, Simpson GL, Solymos P, Henry M, Stevens H, Szoecs E, Wagner H (2017) vegan: Community ecology package. https://cran.r-project.org/web/packages/vegan/index.html
65. Gupta SK, Padmanabhan BR, Diene SM, Lopez-Rojas R, Kempf M, Landraud L, Rolain JM (2014) ARG-ANNOT, a new bioinformatic tool to discover antibiotic resistance genes in bacterial genomes. Antimicrob Agents Chemother 58(1):212–220
66. Austin B, Zhang XH (2006) *Vibrio harveyi*: a significant pathogen of marine vertebrates and invertebrates. Lett Appl Microbiol 43(2):119–124. https://doi.org/10.1111/j.1472-765X.2006.01989.x
67. Rouphael NG, Stephens DS (2012) *Neisseria meningitidis*: biology, microbiology, and epidemiology. Methods Mol Biol 799:1–20. https://doi.org/10.1007/978-1-61779-346-2_1
68. Xu J, Zheng HJ, Liu L, Pan ZC, Prior P, Tang B, Xu JS, Zhang H, Tian Q, Zhang LQ, Feng J (2011) Complete genome sequence of the plant pathogen *Ralstonia solanacearum* strain Po82. J Bacteriol 193(16):4261–4262. https://doi.org/10.1128/Jb.05384-11
69. Ecker DJ, Sampath R, Willett P, Wyatt JR, Samant V, Massire C, Hall TA, Hari K, McNeil JA, Buchen-Osmond C, Budowle B (2005) The microbial rosetta stone database: a compilation of global and emerging infectious microorganisms and bioterrorist threat agents. BMC Microbiol 5:19. https://doi.org/10.1186/1471-2180-5-19
70. Gill EE, Brinkman FSL (2011) The proportional lack of archaeal pathogens: do viruses/phages hold the key? BioEssays 33(4):248–254. https://doi.org/10.1002/bies.201000091
71. Fierer N, Bradford MA, Jackson RB (2007) Toward an ecological classification of soil bacteria. Ecology 88(6):1354–1364. https://doi.org/10.1890/05-1839
72. Roesch LF, Fulthorpe RR, Riva A, Casella G, Hadwin AKM, Kent AD, Daroub SH, Camargo FAO, Farmerie WG, Triplett EW (2007) Pyrosequencing enumerates and contrasts soil microbial diversity. ISME J 1(4):283–290. https://doi.org/10.1038/ismej.2007.53
73. Apweiler R, Bairoch A, Wu CH, Barker WC, Boeckmann B, Ferro S, Gasteiger E, Huang HZ, Lopez R, Magrane M, Martin MJ, Natale DA, O'Donovan C, Redaschi N, Yeh LSL (2004) UniProt: the universal protein knowledgebase. Nucleic Acids Res 32:D115–D119. https://doi.org/10.1093/nar/gkh131
74. Cirillo SLG, Bermudez LE, El-Etr SH, Duhamel GE, Cirillo JD (2001) *Legionella pneumophila* entry gene rtxA is involved in virulence. Infect Immun 69(1):508–517. https://doi.org/10.1128/Iai.69.1.508-517.2001
75. Lin W, Fullner KJ, Clayton R, Sexton JA, Rogers MB, Calia KE, Calderwood SB, Fraser C, Mekalanos JJ (1999) Identification of a *Vibrio cholerae* RTX toxin gene cluster that is tightly linked to the cholera toxin prophage. Proc Natl Acad Sci U S A 96(3):1071–1076. https://doi.org/10.1073/pnas.96.3.1071
76. Suarez G, Khajanchi BK, Sierra JC, Erova TE, Sha J, Chopra AK (2012) Actin cross-linking domain of *Aeromonas hydrophila* repeat in toxin A (RtxA) induces host cell rounding and apoptosis. Gene 506(2):369–376. https://doi.org/10.1016/j.gene.2012.07.012
77. Terada LS, Johansen KA, Nowbar S, Vasil AI, Vasil ML (1999) *Pseudomonas aeruginosa* hemolytic phospholipase C suppresses neutrophil respiratory burst activity. Infect Immun 67(5):2371–2376
78. Aguilera S, De la Torre-Zavala S, Hernández-Flores JL, Murillo J, Bravo J, Alvarez-Morales A (2012) Expression of the gene for resistance to phaseolotoxin (argK) depends on the activity of genes phtABC in *Pseudomonas syringae* pv. phaseolicola. PLoS ONE 7(10):e46815–e46815. https://doi.org/10.1371/journal.pone.0046815
79. Velayati AA, Farnia P, Mirsaeidi M (2015) Persistence of *Mycobacterium tuberculosis* in environmental samples. Int J Mycobacteriol 4:1. https://doi.org/10.1016/j.ijmyco.2014.11.005
80. Leppla SH (1982) Anthrax toxin edema factor—a bacterial adenylate-cyclase that increases cyclic-amp concentrations in eukaryotic cells. Proc Natl Acad Sci Biol 79(10):3162–3166. https://doi.org/10.1073/pnas.79.10.3162
81. Qiu Z (2015) Current pollution status of China's lakes. Paper presented at the The 5th Forum for China Lakes, Jilin
82. Liu L, Johnson HL, Cousens S, Perin J, Scott S, Lawn JE, Rudan I, Campbell H, Cibulskis R, Li MY, Mathers C, Black RE, Who Unicef (2012) Global, regional, and national causes of child mortality: an updated systematic analysis for 2010 with time trends since 2000. Lancet 379(9832):2151–2161. https://doi.org/10.1016/S0140-6736(12)60560-1
83. Li X (2019) Technical solutions for the safe utilization of heavy metal-contaminated farmland in China: a critical review. Land Degrad Dev. https://doi.org/10.1002/ldr.3309

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Submit your manuscript to a SpringerOpen journal and benefit from:

- ► Convenient online submission
- ► Rigorous peer review
- ► Open access: articles freely available online
- ► High visibility within the field
- ► Retaining the copyright to your article

Submit your next manuscript at ► springeropen.com