DBDiaSNP: An Open-Source Knowledgebase of Genetic Polymorphisms and Resistance Genes Related to Diarrheal Pathogens

Kusum Mehla and Jayashree Ramana

Abstract

Diarrhea is a highly common infection among children, responsible for significant morbidity and mortality rate worldwide. After pneumonia, diarrhea remains the second leading cause of neonatal deaths. Numerous viral, bacterial, and parasitic enteric pathogens are associated with diarrhea. With increasing antibiotic resistance among enteric pathogens, there is an urgent need for global surveillance of the mutations and resistance genes primarily responsible for resistance to antibiotic treatment. Single Nucleotide Polymorphisms are important in this regard as they have a vast potential to be utilized as molecular diagnostics for gene-disease or pharmacogenomics association studies linking genotype to phenotype. DBDiaSNP is a comprehensive repository of mutations and resistance genes among various diarrheal pathogens and hosts to advance breakthroughs that will find applications from development of sequence-based diagnostic tools to drug discovery. It contains information about 946 mutations and 326 resistance genes compiled from literature and various web resources. As of March 2015, it houses various pathogen genes and the mutations responsible for antibiotic resistance. The pathogens include, for example, DEC (Diarrheagenic *E.coli*), *Salmonella spp.*, *Campylobacter spp.*, *Shigella spp.*, *Clostridium difficile*, *Aeromonas spp.*, *Helicobacter pylori*, *Entamoeba histolytica*, *Vibrio cholera*, and viruses. It also includes mutations from hosts (e.g., humans, pigs, others) that render them either susceptible or resistant to a certain type of diarrhea. DBDiaSNP is therefore intended as an integrated open access database for researchers and clinicians working on diarrheal diseases. Additionally, we note that the DBDiaSNP is one of the first antibiotic resistance databases for the diarrheal pathogens covering mutations and resistance genes that have clinical relevance from a broad range of pathogens and hosts. For future translational research involving integrative biology and global health, the database offers veritable potentials, particularly for developing countries and worldwide monitoring and personalized effective treatment of pathogens associated with diarrhea. The database is accessible on the public domain at <http://www.juit.ac.in/attachments/dbdiasnp/>.

Introduction

DIARRHEA IS A MANIFESTATION of intestinal dysfunction that results in increased stool output, resulting in loss of water, electrolytes, and/or nutrients. Diarrheal diseases continue to be a significant cause of morbidity and mortality, especially among children under 5 years in developing countries. Over 2.5 million diarrhea-associated deaths were estimated annually concentrated in the most impoverished areas of the world (Baldi, 2009). Oral Rehydration Therapy (ORT) is the cornerstone of treatment as recommended by the WHO (Atia and Buchman, 2009). Addition of zinc to oral

rehydration solution has been proven effective in developing countries (UNICEF/WHO, 2009). However, no evidence suggests efficacy in children living in developed countries. ORT is recommended for all age groups and for acute diarrhea caused by any etiologic agent, but is not sufficient therapy for some cases of bloody diarrhea, severe dehydration, intractable vomiting, and monosaccharide malabsorption.

Continuing increase in antimicrobial resistance among enteric pathogens is becoming a critical area of concern. Antimicrobial resistance, initially a problem of hospital settings associated with large number of nosocomial infections in immune-compromised patients, has now extended itself

to community-acquired infections (WHO, 2012). Emergence of antimicrobial resistance, especially from the multidrug-resistant microbes, is challenging global prospects for fighting the diarrheal pathogens. The emergence and spread of drug-resistant pathogens far outpaces the development of replacement drugs, shrinking the therapeutic arsenal (Dantas and Sommer, 2014).

The antibiotic resistance mechanisms are diverse and complex. Bacteria have developed resistance to all different classes of antibiotics discovered so far and appear to justify the ‘Dawn of Post-Antibiotic Era’. Acquired resistance that results from mutation or acquisition of foreign resistance genes is a major player in the resistance against antibiotics (Harbottle et al., 2006). Antibiotic resistance is the evolutionary consequence of deployment of a selective pressure. Mutations and the resistance genes in the pathogen genome affect the drug response by various mechanisms such as inactivation of antimicrobial agents, modification of the antimicrobial target, altering membrane permeability, and efflux pumps (Aleksun and Levy, 2007), making it crucial to understand this trilateral relationship between disease and genome variations and resistance genes.

Single Nucleotide Polymorphisms (SNPs) are the unique genetic differences between individuals that contribute significantly in identification and forensics (Sobrinho et al., 2005), mapping and genome-wide association studies to complex diseases (Cai et al., 2013), predicting specific genetic traits (Yao et al., 2014), and classifying patients in clinical trials (Pinto et al., 2014). SNPs underlie differences in susceptibility to or protection from a host of diseases, thereby proving its great potential in personalized medicine (Chakravarti, 2001). Therefore, a database of the variations in a group of organisms is not only useful for understanding genotype–phenotype relationships but also in clinical applications. The phenomenon of gene acquisition is mediated by transformation, transduction, and conjugation. Plasmids, transposons, integrons, and gene cassettes are the platforms on which these resistance genes are assembled. They allow bacterial strains to expand their niches to the areas that were previously denied (Bennett, 2008).

Large-scale studies have been undertaken throughout the world to identify the antibiotic resistance-associated mutations that are devoted to some specific organism such as *Mycobacterium tuberculosis* (MTCID, MGDD, MTBC,

FIG. 1. DBDiaSNP architecture overview. **(a)** Standard three-tier architecture of DBDiaSNP; **(b)** Organizational structure of the DBDiaSNP. The data were compiled from existing literature and online databases Uniprot, ClinVar, dbSNP, and OMIM, and enriched with annotation information such as PDB Id, Gene Id, Pubmed Id, Biological function, and CCDS Id from online resources. The database includes three modules: 1) Polymorphism data further categorized into Host and Pathogen Polymorphism; 2) Resistance Gene data further categorized into Plasmid mediated, Integrons, and Gene Cassettes, and Transposon Mediated; 3) Research articles relevant to each category.

GMTV) or which are generalized and hosts mutations or resistance data from different organisms (CRAD, ARDB, CREAM) (Baux et al., 2008; Bharti et al., 2012; Chernyaeva et al., 2014; Dai et al., 2014; Liu and Pop, 2009; McArthur et al., 2013; Stucki and Gagneux, 2013; Vishnoi et al., 2008). The information about pathogens that are implicated in other disorders, including diarrhea, is not disease specific, while our database pertains to clinical isolates of diarrheal pathogens.

To date, there has been no single resource available that provides information about genomic variations and resistance genes among different diarrheal pathogens, readily available to be used in research and development. We have developed DBDiaSNP in an attempt to provide a single platform for storage and retrieval of a large amount of polymorphism and resistance gene data, particularly that derived from clinical isolates of diarrhea. It complements our other database dbDiarrhea (Ramana and Tamanna, 2012), which is a repertoire of pathogen proteins and vaccine antigens implicated in the pathogenesis of diarrhea. This database provides information about prevalence of resistance genes or mutations in a specific geographical region, increasing our knowledge on which mutations or resistance genes are common in certain bacteria or geographical locations. Mutation data in such databases can be used to screen newly sequenced strains for presence of any previously characterized resistance associated mutation, ultimately paving a way for information based targeted therapies. By mapping these SNPs at a structural level, we can glean insights into the molecular level details of antibiotic resistance due to mutations. We can study how these mutations affect the binding of antibiotics to the target proteins. Such studies offer us with the possibility of mitigating the ever growing problem of antibiotic resistance by incorporating necessary modifications in already existing antibiotics.

Methods

Database System Implementation

DBDiaSNP has been designed on MySQL 5.1 that provides multi-layered server design with fast execution times. It is implemented by using three-tier architecture. DBDiaSNP was developed on an IIS server (7.0) in a Windows operating system. The web pages were written using PHP, html language, Cascading Style Sheets (CSS), and JavaScript. The IIS server handles queries from web clients through PHP scripts to perform searches, which allow researchers to obtain SNP information. An overview of the DBDiaSNP architecture is given in Figure 1a.

As shown in Figure 1b, the database is a comprehensive resource of drug-resistant mutations and resistance genes in diarrheal pathogens and hosts. We conducted a systematic review to identify antibiotic resistance mutations and resistance genes from literature to include in the database. The variations are categorized into SNPs, substitutions, frame-shift mutations, deletions, and insertions. For each mutation, the database provides complete codon change at nucleotide or amino acid level. Resistance genes reported in different pathogens can also be browsed and are categorized into Plasmid mediated, Transposons mediated, Integrins, and gene Cassettes mediated resistance modes according to

TABLE 1. SNPs STATISTICS: SPECIES DISTRIBUTION

Organism	# SNPs
<i>Host polymorphism</i>	
Porcine	65
Humans	291
<i>Pathogen polymorphism</i>	
Viruses	221
<i>Escherichia coli</i>	91
<i>Entamoeba histolytica</i>	11
<i>Campylobacter spp.</i>	98
<i>Shigella spp.</i>	13
<i>Vibrio cholera</i>	16
<i>Aeromonas caviae</i>	4
<i>Clostridium difficile</i>	46
<i>Helicobacter pylori</i>	71
<i>Salmonella spp.</i>	19
Total	946

subcellular locations. Table 1 and Table 2 show the database statistics that houses 946 mutations with clinical relevance and 326 resistance genes associated with diarrheal pathogens (viruses, protozoans, and bacterial species) and hosts. In pathogens, mutations and resistance genes are responsible for antibiotic resistance. The host mutations are the ones that render them susceptible or resistant to a particular type of diarrhea or to diseases where diarrhea is a predominant symptom.

SNP and Resistance Data Search

DBDiaSNP allows users to search for SNP and resistance gene information based on organism of interest that includes both host and pathogens. Related literature citations can also be retrieved based on the resistance mechanism adopted. The search results provide related SNP and resistance gene data and their annotations. The information on corresponding protein, PDB identifiers, and functions associated with each of the SNPs and resistance genes has been extracted from the web resources dbSNP, OMIM, ClinVar, and PDB (Berman et al., 2000; Hamosh et al., 2005; Landrum et al., 2013; Sherry et al., 2001), thus providing a comprehensive compilation of the SNPs and resistance genes reported in the clinical isolates of various diarrheal pathogens.

TABLE 2. RESISTANCE GENE STATISTICS: SPECIES DISTRIBUTION

Organism	# Resistance genes
<i>Escherichia coli</i>	68
<i>Enterococcus spp.</i>	10
<i>Campylobacter spp.</i>	10
<i>Shigella spp.</i>	58
<i>Vibrio spp.</i>	60
<i>Aeromonas spp.</i>	26
<i>Clostridium spp.</i>	10
<i>Klebsiella spp.</i>	16
<i>Salmonella spp.</i>	68
Total	326

FIG. 2. Typical output of a query (SNP). (a) Description of different annotation terms; (b) Distribution of mutations associated with antibiotic resistance in *C. difficile*; (c) Mutation in IL8 associated with susceptibility to EAEC-associated traveler's diarrhea in *H. sapiens*.

Results and Discussion

DBDiaSNP presents a database entirely devoted to drug-resistance mutations and resistance genes in diarrhea. By providing a comprehensive repository of drug resistance mutation and resistance genes, we hope it will open new avenues to the discoveries that will have applications from diagnostics to drug discovery.

Development of SNP-based diagnostics

There is pressing need to develop tools that can identify the antibiotic-resistant pathogens rapidly and with high accuracy. SNPs-based diagnostic methods detect mutations associated with a particular disease, but such tools extensively rely on information about the mutations and their relative frequency (Palzkill, 2013). As more sequencing data are made available, this database project, we believe, will pave a path for the currently available tools to detect resistance to the antibiotics.

SNPs monitoring in different geographical regions

Different analogs of the same drug interact differently with the targets. Mutations arising as a result of treatment with different drug analogs may help explicate the geographic differences in drug effectiveness (Bennett et al., 2009). Therefore, we have included the geographic locations of antibiotic resistance studies in our database. As more data will be available on frequency of mutations through sequencing projects, we will have a better understanding of significance of regional differences in drug resistance.

Drug discovery

SNPs can have a major influence on response to pharmacotherapy (Voisey and Morris, 2008). The database will aid in discovery of antibiotic-resistant pathogens by mapping the mutations onto the protein structure, and providing a better understanding of the molecular mechanism of drug resistance. Additionally, such analysis may suggest drug modifications to combat emerging drug resistance in diarrheal pathogens.

We have created an interactive web interface to retrieve information about mutations and resistance genes in diarrheal pathogens associated with drug resistance. Currently there is information from major diarrheal pathogens and hosts. To get a comprehensive list of variations, particularly substitutions and in-dels, users can choose Host Polymorphism or Pathogen Polymorphisms. As shown in Figure 2 for both pathogen and host mutations, after submission, the output will show all the SNPs along with annotation of genes corresponding to these SNPs. Users can also browse through the resistance genes in different diarrheal pathogens. Figure 3 shows the results of a query executed with *Campylobacter spp.* listing all the resistance genes and related annotation data. Users can also search for the related literature based on categories polymorphism, plasmid-mediated resistance, or integrons and gene-cassettes.

In addition to presenting the literature in a manually curated database, this web interface will also serve as a gateway to post data from future research. In keeping with the scientific philosophy of open access, the database is publicly accessible. We hope that this database will serve as a portal to share future research on antibiotic resistance.

Campylobacter Spp. <input type="button" value="Retrieve"/>									
Gene_Name	Protein_Name	Location	Organism	Geographical_Location	Resistance_Mechanism	Clinical_Significance	Pubmed_Id	Uniprot_Id	GenBank
aadA	Aminoglycoside adenyltransferase	Plasmid	Campylobacter spp.	France	Plasmid Mediated	confers resistance to streptomycin	2168151	Q8KRK6	
aadE	Streptomycin aminoglycoside 6-adenyltransferase	Plasmid	Campylobacter spp.	France	Plasmid Mediated	confers resistance to streptomycin and spectinomycin	2168151	Q4VRA5	
ant(6)-Ib	Streptomycin aminoglycoside adenyltransferase ant(6)-Ib	Plasmid	Campylobacter fetus subsp. Fetus	Switzerland	Plasmid Mediated	confers resistance to streptomycin	20479200	D8L9Z0	
cmeB	RND efflux system, inner membrane transporter CmeB	Efflux Pump (cmeABC)	Campylobacter spp.	United States	Efflux pump	resistance to ciprofloxacin, erythromycin, and tetracycline	16048946	Q5HWA1	321
cmeC	Outer membrane channel protein CmeC	Efflux Pump (cmeABC)	Campylobacter jejuni	United States	Efflux pump	confers resistance to cefotaxime and erythromycin	15728904	Q0PBE5	90
cmeR	Transcriptional repressor	Efflux Pump (cmeABC)	Campylobacter jejuni	United States	Efflux pump	confers resistance to cefotaxime and erythromycin	15728904	Q7B8P6	46
dfi1	Dihydrofolate reductase	Integron 1	Campylobacter jejuni	Sweden	Integrons and Gene Cassettes	confers resistance to trimethoprim	10990262	Q9EXQ6	
OXA-61	OXA-61 beta-	Plasmid	Campylobacter	Australia	Plasmid Mediated	confers resistance to	15917560		

FIG. 3. Typical output of a query (Resistance Gene). This view shows the distribution of resistance genes distributed in *Campylobacter spp.*

Future directions

The current version of the database is based on literature available up to March 2015. The database will be updated regularly to provide an accurate picture of the mutations and resistance genes associated with antibiotic resistance in diarrheal pathogens. To keep the database updated, we will include new mutation prevalence data sets. We will include novel mutations from published studies in clinical isolates of drug-resistant diarrheal pathogens. We will customize the website to house more plasmid-mediated resistance and integrons and gene cassette-associated resistance mechanisms and related mutations in the near future.

Conclusions

In this article we have described DBDiaSNP, a comprehensive database of molecular variations and resistance genes in clinical isolates of diarrhea, generated through manual curation of data from published literature. DBDiaSNP is intended to pace up with research efforts for global surveillance and control of diarrhea. The variations reported in the database represent potential markers of antibiotic resistance in diarrheal pathogens that might help in understanding genotype-phenotype relationships. The database will expand as additional polymorphisms will be identified in the coming years. We expect that it will make a substantial contribution towards the breakthroughs that will find applications ranging from diagnostics to drug discovery and will prove to be useful to both the diarrheal research community and clinicians. For future translational research involving integrative biology and global health (Ahmed et al., 2014; Dandara et al., 2014), the presented database offers veritable potentials, particularly for developing countries and worldwide monitoring and personalized effective treatment of pathogens associated with diarrhea.

Acknowledgments

This research was supported by DST (Department of Science and Technology), Ministry of Science and Technology, India under Grant Number SB/FT/LS-278/2012. We thank Tamanna and Ashwani Kumar for their long-standing support and invaluable suggestions. The authors would like to thank the anonymous reviewers for their insightful comments that helped us improve the earlier versions of the manuscript.

Author Disclosure Statement

We confirm that there are no conflicts of financial interest associated with this publication.

References

- Ahmed AA, Mohamed AA, Guled IA, et al. (2014). Knowledge translation in Africa for 21st century integrative biology: The “know-do gap” in family planning with contraceptive use among Somali women. *OMICS* 18, 696–704.
- Alekshun MN, and Levy SB. (2007). Molecular mechanisms of antibacterial multidrug resistance. *Cell* 128, 1037–1050.
- Atia AN, and Buchman AL. (2009). Oral rehydration solutions in non-cholera diarrhea: A review. *Am J Gastroenterol* 104, 2596–2604.
- Baldi F. (2009). Focus on acute diarrhoeal disease. *World J Gastroenterol* 15, 3341–3348.
- Baux D, Faugere V, Larrieu L, et al. (2008). UMD-USHbases: A comprehensive set of databases to record and analyse pathogenic mutations and unclassified variants in seven Usher syndrome causing genes. *Human Mutat* 29, E76–87.
- Bennett D, Camacho R, Otelea D, et al., (2009). Drug resistance mutations for surveillance of transmitted HIV-1 drug-resistance: 2009 Update. *PLoS One* 4, e4724.
- Bennett PM. (2008). Plasmid encoded antibiotic resistance: Acquisition and transfer of antibiotic resistance genes in bacteria. *Br J Pharmacol* 153, S347–S357.
- Berman H, Westbrook J, Feng Z, et al. (2000). The Protein Data Bank. *Nucleic Acids Res* 28, 235–242.
- Bharti R, Das R, Sharma P, et al. (2012). MTCID: A database of genetic polymorphisms in clinical isolates of *Mycobacterium tuberculosis*. *Tuberculosis* 92, 166–172.
- Cai M, Dai H, Qiu Y, et al. (2013). SNP set association analysis for genome-wide association studies. *PLoS One* 8, e62495.
- Chakravarti A. (2001). Single nucleotide polymorphisms: To a future of genetic medicine. *Nature* 409, 822–823.
- Chernyaeva EN, Shulgina M, Rotkevich M, et al. (2014). Genome-wide *Mycobacterium tuberculosis* variation (GMTV) database: A new tool for integrating sequence variations and epidemiology. *BMC Genomics* 15, 308.
- Dai E, Lv Y, Meng F, et al. (2014). CREAM: A database for chemotherapy resistance-associated miRSNP. *Cell Death Dis* 5, e1272.
- Dandara C, Huzair F, Borda-Rodriguez A, et al. (2014). H3Africa and the African life sciences ecosystem: Building sustainable innovation. *OMICS* 18, 733–739.
- Dantas G, and Sommer MOA. (2014). How to fight back against antibiotic resistance. *Am Sci* 102, 42–51.
- Hamosh A, Scott A, Amberger J, et al. (2005). Online Mendelian inheritance in man (OMIM), a knowledgebase of human genes and genetic disorders. *Nucleic Acids Res* 33, D514–D517.
- Harbottle H, Thakur S, Zhao S, et al. (2006). Genetics of antimicrobial resistance. *Animal Biotechnol* 17, 111–124.
- Landrum M, Lee J, Riley G, et al. (2013). ClinVar: Public archive of relationships among sequence variation and human phenotype. *Nucleic Acids Res* 42, D980–985.
- Liu B, and Pop M. (2009). ARDB—Antibiotic Resistance Genes Database. *Nucleic Acids Res* 37, D443–447.
- McArthur A, Wagelchner N, Nizam F, et al. (2013). The comprehensive antibiotic resistance database. *Antimicrob Agents Chemother* 57, 3348–3357.
- Palzkill T. (2013). Mitigating antibiotic resistance with DNA sequence information. In: *Emerging and Persistent Infectious Diseases: Focus on Antimicrobial Drug Resistance*. pp. 19–22.
- Pinto N, Gamazon E, Antao N, et al. (2014). Integrating cell-based and clinical genome-wide studies to identify genetic variants contributing to treatment failure in neuroblastoma patients. *Clin Pharmacol Therapeut* 95, 644–652.
- Ramana J, and Tamanna. (2012). dbDiarrhea: The database of pathogen proteins and vaccine antigens from diarrheal pathogens. *Infect Genet Evol* 12, 1647–1651.
- Sherry S, Ward M, Kholodov M, et al. (2001). dbSNP: The NCBI database of genetic variation. *Nucleic Acids Res* 29, 308–311.
- Sobrinho B, Brión M, and Carracedo A. (2005). SNPs in forensic genetics: A review on SNP typing methodologies. *Foren Sci Intl* 154, 181–194.

- Stucki D, and Gagneux S. (2013). Single nucleotide polymorphisms in *Mycobacterium tuberculosis* and the need for a curated database. *Tuberculosis* 93, 30–39.
- UNICEF/WHO. (2009). WHO | Diarrhoea: Why children are still dying and what can be done. World Health Organization.
- Vishnoi A, Srivastava A, Roy R, et al. (2008). MGDD: *Mycobacterium tuberculosis* genome divergence database. *BMC Genom* 9, 373.
- Voisey J, and Morris CP. (2008). SNP technologies for drug discovery: A current review. *Curr Drug Disc Technol* 5, 230–235.
- WHO. (2012). WHO | The evolving threat of antimicrobial resistance—Options for action. World Health Organization.
- Yao C, Leng N, Weigel KA, et al. (2014). Prediction of genetic contributions to complex traits using whole genome sequencing data. *BMC Proc* 8, S68.

Address correspondence to:

Jayashree Ramana, PhD

Department of Biotechnology and Bioinformatics

Jaypee University of Information Technology

Solan

PIN-173234 Himachal Pradesh

India

E-mail: jayashree.ramana@juit.ac.in