Genetic diversity of pathogenic microorganisms: Basic insights, public health implications and the Indian initiatives

J. S. Virdi* and Pooja Sachdeva

Microbial Pathogenicity Laboratory, Department of Microbiology, University of Delhi South Campus, Benito Juarez Road, New Delhi 110 021, India

Although pathogenic microorganisms constitute a small proportion of the microbial species, these are characterized by high genetic diversity. In the past decade or so, there has been considerable advancement in understanding the genetic population structures of pathogenic microorganisms that may range from strictly clonal to effectively panmictic. We, now understand that besides the staple mechanisms like point mutations, gene duplication, gene rearrangements and gene loss, horizontal (lateral) gene transfer including pathogenicity islands play an important role in the generation of genetic diversity among microbial pathogens. Understanding genetic diversity of pathogens may have far-reaching implications for public health intervention strategies such as tracking the global spread of pathogens, understanding emergence of new and drug-resistant microbes and rational development of diagnostics, therapeutics and vaccines. The study of pathogen diversity may also help us better understand evolution, taxonomy and pathogenicity *per se*. The endeavours to understand genetic diversity of extant pathogenic microorganisms in India are discussed here.

PATHOGENIC microbes, though constitute a very small proportion of the microbial species, are nevertheless characterized by high genetic diversity[1–3]. Genotypic variation in pathogen populations poses a major barrier to disease control. As of today, amongst the pathogenic microorganisms, diversity of bacterial pathogens infecting human or animal hosts is the most studied. Considerably less is known about the genetic diversity of plant pathogens, though it is quite understandable that basic concepts would be the same as applicable to human pathogens. Other pathogenic microbial forms, viz. viruses, disease-causing protozoa and pathogenic fungi would, however, have their own peculiarities in respect of diversity besides the common denominators.

For any biodiversity study, an appropriate sample of the population is a prerequisite. In contrast to the diversity of non-pathogenic microbes, for which study of reasonably good number of samples collected from diverse terrestrial and aquatic habitats would suffice, study of the diversity of microbial pathogens is dependent largely on isolates deposited with culture collection centers. The culture collection centers however do not represent actual status of diversity extant in nature, as majority of the pathogenic strains in such collections are obtained from clinical settings and thus, biased towards the most pathogenic forms with the exclusion of opportunistic and accidental pathogenic forms[4]. The same is true for bacteria that are primarily environmental such as *Campylobacter jejuni*. Culture collections in such cases mainly comprise isolates that cause frank human disease, and consequently do not represent the actual pathogen diversity. Thus, among the impediments in our understanding of pathogen diversity, problems associated with accurate sampling remain paramount. These must be overcome by undertaking structured sampling programmes, rather than simply storing the isolates, which happen to reach the scientist through diagnostic or similar laboratories[5]. For obligate human pathogens such as *Mycobacterium tuberculosis*, which are known not to have a reservoir elsewhere than human subjects, culture collections however represent a reasonably accurate sample of pathogen population. Another problem associated with studying pathogen diversity is selection pressure of the immune system of the host, which imposes a bias by perpetually selecting certain genotypes.

The various methods which have been used to study genetic diversity in microorganisms including pathogens are – restriction analysis of chromosomal DNA (REAC), restriction fragment length polymorphism (RFLP), randomly amplified polymorphic DNA (RAPD), pulsed-field gel electrophoresis (PFGE), amplified fragment length polymorphism (AFLP), variable number of tandem repeats (VNTR), repetitive sequence element-based strategies like REP (repetitive extragenic palindrome)-PCR, ERIC (enterobacterial repetitive intergenic consensus)-PCR and BOX-PCR based fingerprinting, multilocus enzyme electrophoresis (MLEE) and, single locus and multilocus sequence typing (MLST) including multi-virulence locus sequence typing (MVLST). An enormous amount of data on the genetic diversity of a number of common pathogens has been generated using these techniques[6]. However, each one of these techniques, except MLST, suffers from one or a combination of limitations like lack of intra- or inter-laboratory reproducibility, tedious methodology,

*For correspondence. (e-mail: virdi_dusc@rediffmail.com)

difficulty in interpretation of results, and non-applicability to certain pathogens. Currently, in spite of the high cost involved, MLST seems to be the only fool proof and fully portable technique for undertaking genetic diversity studies[4]. With the high throughput DNA sequencing becoming more and more economical and within reach of most laboratories, and the availability of high throughput high density resequencing microarrays[7], the area of genetic diversity is sure to witness a radical change.

Besides the problems associated with accurate sampling and lack of availability of an economical fool proof genotyping method, lack of good framework for interpretation of genetic diversity data is another stumbling block. This article critically analyses various aspects of the intraspecific genetic diversity of bacterial pathogens. These include – clonality of pathogens, principles underlying generation of genetic diversity including horizontal gene transfer and pathogenicity islands, public health implications of diversity such as emergence of new and drug-resistant microbes, and the long-term benefits of studying diversity in understanding evolution of pathogens and pathogenicity *per se*.

Basic concepts and insights

Clonality paradigm and panmixia

Prior to the advent of molecular biology techniques, the tool, which has been widely used for studying genetic diversity in microbes, is multilocus enzyme electrophoresis (MLEE or MEE). Multilocus enzyme electrophoresis indexes allelic variations of several structural genes and thus helps in estimating the overall genotypic diversity in the species[8]. Basic metabolic enzymes are analysed, which are expressed in all isolates of a species. Such allelic variations in bacteria are unaffected by environmental factors and are minimally subject to convergent evolution[9,10]. Moreover, it has been shown that dendrograms generated by MLEE are generally concordant with phylogenetic trees based on extensive nucleotide sequence analyses[11–13]. Extensive applications of MLEE to study genotypic diversity of *E. coli* in the past indicated strong linkage disequilibrium, i.e. non-random association of alleles, which led to the assumption that *E. coli* was clonal.

First recognized in *E. coli*, the clone concept was later elevated to a paradigm extending it to all populations of bacteria. Other evidences that have generally been taken to construe clonal nature of bacteria are: asexual mode of reproduction with many species exhibiting strong linkage disequilibrium between different loci, frequent recovery of only a few of all the possible multilocus genotypes, association of only particular serotypes with disease[14], and repeated recovery of isolates having identical multilocus genotypes globally or from widely separated geographical regions. Extensive studies with other species however revealed that bacteria were not invariably clonal[15]. The high level of allelic diversity at enzyme loci in most species of bacteria shows that virtually unlimited number of distinctive multilocus genotypes is potentially possible. Yet, for a majority of the pathogenic species studied, the actual number of clones recovered is very small, indicating that, in general, genotypic diversity in pathogenic species is much lower than the non-pathogenic forms. Does that mean that pathogenic bacteria are clonal? Absolutely not. A case in point is that of *Neisseria gonorrhoeae*[16], the causative agent of gonorrhea in which gene exchange occurs at random, leading to highly panmictic population. Even the limited genetic variability exhibited by *Neisseria meningitidis* is not true clonality. The apparent clonality of this pathogen is based on an 'epidemic clone' or electrophoretic type (ET5), which appeared in early seventies and has since then, caused epidemics in several countries. Based on statistical testing, it was argued that due to frequent recombination, such clones would disappear in future. In fact, the variant clones of *N. meningitidis* ET5 that differ at one or more loci are already becoming increasingly common[17]. Among pathogenic bacteria, true clonality (long-term clonal evolution) is best exemplified by *Salmonella enterica*[18] and *Borrelia burgdorferi*[19]. Thus, the clonal model has been modified and it is now established that population structure of bacteria ranges from effectively panmictic to strictly clonal.

Like bacteria, the genetic populations of pathogenic protozoa and fungi too range from clonal to panmictic. In fact, M. Tibayrenc who has extensively studied the genetic diversity of pathogenic protozoa recommends a unified approach for studying diversity of all pathogenic microbes namely bacteria, fungi and protozoa[20]. Studies from his laboratory have sufficiently argued for the clonal population structure of several parasitic protozoa, viz. *Entamoeba histolytica*, *Giardia duodenalis*, *Leishmania* spp., *Trypanosoma brucei* and *T. cruzi* both by linkage disequilibrium studies and deviation from Hardy–Weinberg expectations[21]. However, leaving aside *T. cruzi* – the causative agent of the Chagas disease – studies have revealed that apparent clonal population structure in other species actually represents an epidemic population resulting from the explosive spread of a single type. Recent studies have revealed that *Plasmodium falciparum* has a panmictic population structure[22].

The genetic diversity of pathogenic fungi is understood the least. *Candida albicans*, one of the most commonly isolated opportunistic organisms, has been studied extensively and results ranging from absence or little linkage disequilibrium to very clear evidences of clonality have been obtained[20]. Another opportunistic yeast, *Cryptococcus neoformans* also exhibits clonal population structure[20]. *Pneumocystis carinii*, a eukaryotic pathogenic microorganism with fungal affinities which causes respiratory infections in immunocompromised patients, has also been shown to be clonal by multilocus enzyme electrophore-

sis[23]. However more studies are required to confirm or refute these observations.

Mechanisms of genetic variability

Although asexual, prokaryotes possess several mechanisms that result in gene shuffling leading to immense genetic diversification. It is said that you name it, and the microbes including the pathogenic microorganisms have it – point mutations, gene duplication, gene silencing, gene rearrangement and gene loss. Even though these staple mechanisms result in only slow and subtle changes, these have important implications with respect to generation of diversity. Random genetic mutations in the pre-existing genes result in functional modification and elimination of original gene(s), contributing to bacterial virulence. Such mutations (pathogenicity adaptive or pathoadaptive mutations) might result in emergence of highly pathogenic bacterial lineages by providing host-specific adaptations to the pathogen. For example, uropathogenic strains of *E. coli* carrying minor mutations in *fim* H gene for the adhesive subunit of type I fimbriae, possess an enhanced ability to recognize mono-mannose receptors and exhibit higher tropism for uroepithelium[24]. Therefore, these mono-mannose specific strains have a higher ability to colonize bladder than those harbouring a wild-type allele, providing a selective advantage in their virulence niche. Point mutations can also generate antigenic diversity, with consequent evasion from immune response of the host, leading to stabilization of the genetic variant.

Loss of resident genes (genomic deletions) in an organism also plays an important role in generation of pathogen diversity. In *Pseudomonas aeruginosa*, deletions in the *muc* A gene result in the overproduction of alginate and thus provide an additional advantage for their survival in the lungs[25,26]. Similarly, loss of lysine decarboxylase gene (*cad* A) through genetic deletions appear to favour *Shigella* and enteroinvasive *E. coli* (EIEC) strains, as the product of this gene inhibits enterotoxin activity. In addition to loss of *cad* A gene, deletion of *omp* T gene for a surface protease that attenuates expression of *vir* G, also increases the pathogenic potential of some of the organisms[27].

Transposable elements like transposons and insertion sequences (IS) are important mobile genetic elements that easily promote transfer of genes between phylogenetically diverse populations contributing to the genetic diversity. In addition to simple transposition, IS elements also give rise to complex DNA rearrangements such as deletions, inversions, gene amplifications and fusion of two DNA molecules by co-integrate formation. IS elements are also passively involved in recombination because the flanking homologous DNA sequences might serve as recombinational cross-over sites. Apparently, IS-mediated mechanisms play a major role in generation of genetic diversity. Transposons are generally involved in propagation of antibiotic-resistant genes, for example transposon Tn5 confers resistance to three antibiotics, viz. kanamycin, bleomycin and streptomycin[28]. In addition to transposons, transfer of genes is also mediated by integrons. Integrons are gene expression elements that incorporate promoterless genes by site-specific recombinational events mediated by integrases. Integrons can incorporate several genes next to its promoter leading to stockpiling of many resistance and virulence genes contributing to enhanced pathogenicity of an organism[29,30]. Recently, integrons have been implicated in the acquisition of virulence determinants in *V. cholerae*[31].

Horizontal gene transfer and pathogenicity islands

Horizontal (lateral) gene transfer (HGT) refers to the genetic exchange among different evolutionary lineages. Until recently, the role and the extent of HGT in bacteria were debatable but development of high-throughput sequencing and comparative genomics has put forth an unprecedented evidence of the presence of this mechanism in bacteria[32]. Ochman and coworkers have recently revealed that the extent of HGT in bacteria varies from virtually none in some bacteria like *Borrelia burgdorferi* to up to 12.8% in *E. coli*[33] (Figure 1). Although the processes underlying HGT, i.e. transformation, transduction and conjugation are common to both pathogenic and non-pathogenic bacteria, their implications are highly variable with respect to pathogenic organisms.

HGT and recombination not only play an important role in major events like emergence of new and highly virulent pathogens as exemplified by enterohaemorrhagic *E. coli* O157 : H7, but also confer novel capabilities on the recipients, allowing them to conquer new virulence niches, leading to diversification of natural populations. The latter is best illustrated by *Streptococcus pyogenes* and penicillin-resistant *Streptococcus pneumoniae*. The presence of an allele (*spe* A1) encoding scarlet fever toxin[12] in diverse phylogenetic lineages of *S. pyogenes* meant that it has been distributed horizontally among clones of this organism. In penicillin-resistant *S. pneumoniae*, HGT and recombinational processes have been shown to generate mosaic genes leading to molecularly remodelled penicillin-binding proteins (PBPs) with reduced affinity to penicillins[34,35], and variations in the genes for capsular antigen[36] and immunoglobulin A1 proteases[37]. Also, multiple episodes of HGT among divergent phylogenetic lineages of *Staphylococcus aureus* seem to have been responsible for the global spread of the methicillin-resistant *S. aureus*[38]. Extensive horizontal genetic exchange was reported in *Campylobacter jejuni* both *in vivo* and *in vitro* and has been thought to underlie the genome plasticity observed in this pathogen[39].

Bacteriophage-mediated transfer of genetic material between organisms is a common phenomenon in several pathogenic bacteria. Certain phages carry virulence genes within their genomes that play an important role in bacte-

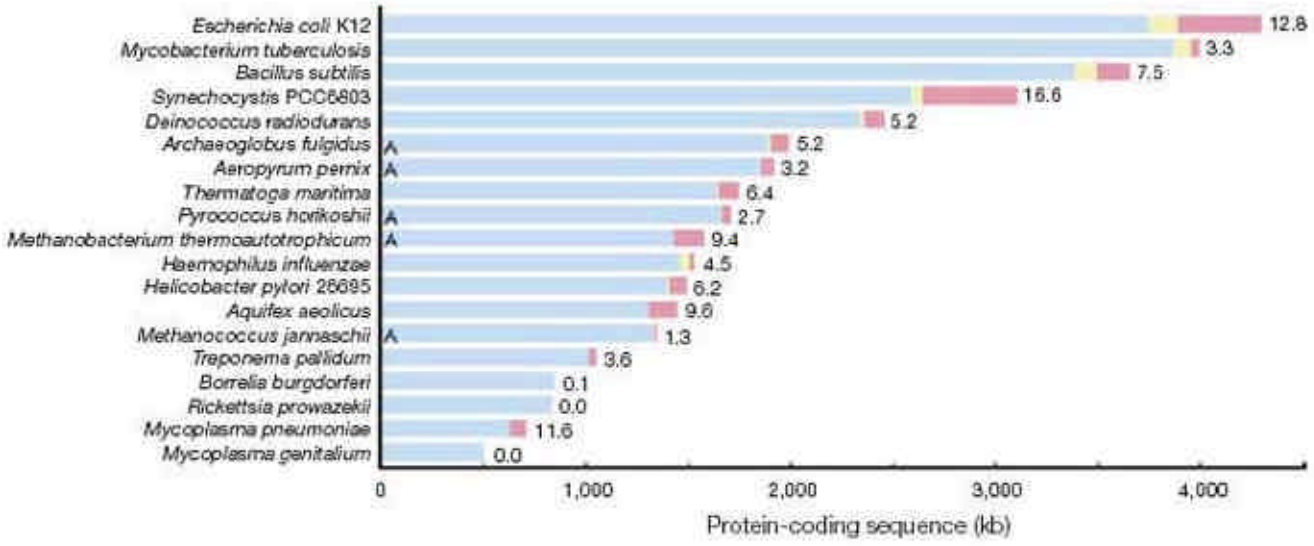


Figure 1. Distribution of horizontally acquired (foreign) DNA in sequenced bacterial genomes. Lengths of bars denote the amount of protein-coding DNA. For each bar, the native DNA is blue; foreign DNA identifiable as mobile elements, including transposons and bacteriophages, is yellow, and other foreign DNA is red. The percentage of foreign DNA is noted to the right of each bar. 'A' denotes an Archaeal genome. (From ref. 33; reproduced with permission.)

rial pathogenesis and upon lysogenization can convert a strain to its 'pathogenic variant'. For example, cholera toxin genes of *Vibrio cholerae* are encoded within the genome of a filamentous phage[40]. In addition to conferring virulence properties on the infected host, bacteriophages can also mediate transfer of large fragments of DNA by generalized transduction, thereby generating genome variability. Staphylococcal phages – $f13$ and $80a$ play an important role in the mobility of pathogenicity island encoding toxic shock toxin-1 in *S. aureus*[41]. Certain giant broad-range phages like fKZ, can mediate exchange of genetic modules within a network of closely and distantly related genomes[42].

Horizontally acquired 'pathogenicity islands' (PAIs) are one of the major contributors to the pathogenic lifestyle of certain bacteria. PAIs are a subset of genomic islands first described in *E. coli* but now identified in genomes of plant, animal and human pathogens encoding large clusters of virulence traits[43]. These PAIs often reside at t-RNA or t-RNA like loci and are flanked by direct repeats mediating their mobility[44–46]. Also, ORFs within certain PAIs show homology to bacteriophage integrase. All these features along with the differences in the G + C content and codon usage of PAIs when compared to the genome in which they reside, indicate that PAIs are acquired by HGT mediated either by phages or conserved integrases. Recently, comparison of genomes of virulent, less virulent and avirulent strains by subtractive hybridization suggested that PAIs have important implications in conferring virulence to pathogenic strains. PAI I (encoding a-hemolysin) and PAI II (encoding a-hemolysin and P-related fimbriae) are two well-characterized pathogenicity islands that play an important role in the pathogenicity of uropathogenic *E. coli*[47]. In enteropathogenic *E. coli* (EPEC), a 35-kb type III secretion (TTS)-related pathogenicity island called the locus of enterocyte effacement (LEE) encodes all the genes required for its attachment and effacing properties. It has been demonstrated that cloned LEE confers these properties to non-pathogenic *E. coli* K12.

Acquisition of PAIs can transform a benign organism to a potential pathogen and can play an important role in the evolution and establishment of new pathogens. This probably has been the case with *Yersinia pestis*, the causative agent of plague. The population genetic structure analysis of *Y. pestis* and two other yersiniae, viz. *Y. pseudotuberculosis* and *Y. enterocolitica* have shown that *Y. pestis* is a clone that has emerged recently from *Y. pseudotuberculosis*[48]. Also, 'high pathogenicity islands' (HPI) of yersiniae presumably served as one of the key factors in the emergence of highly virulent *Y. pestis* from *Y. pseudotuberculosis*[44].

Public health implications

Although genetic diversity allows pathogenic microorganisms to exploit diverse hosts and environmental niches, escape the onslaught of the immune response and help in evolution, it can also be exploited for devising public health intervention programmes. These include epidemiological investigations aimed at tracking the origin and spread of disease-causing organisms, and rational development of diagnostics, therapeutics and vaccines.

Molecular epidemiology

Nowhere else has the knowledge of genetic diversity of pathogens exploited as directly as in the study of epide-

miology. In the molecular epidemiology parlance, this is commonly referred to as 'typing' or 'genotyping'. It is being practised widely in medical microbiology and has proved an invaluable tool in tracking strains responsible for disease outbreaks. This has particularly been useful in studying and controlling nosocomial outbreaks. It has also been used to study whether the relapse of an infectious disease after therapeutic intervention, was due to treatment failure or recolonization of the host by a new strain. Equally important is the study of global epidemiology of pathogens wherein long-term movements of disease-causing microbes are followed over wide geographical areas. Studying global epidemiology of a pathogen is a complex task and cannot be accomplished without the availability of an exhaustive collection of genetic variants obtained from different regions of the world.

Diagnostics and therapeutics

It is well established that a particular disease may be caused by closely related but distinct genetic variants of a pathogen. Poorly understood diversity of a pathogen can lead to the development of diagnostics with pitfalls in detecting the entire set of pathogen genotypes that can cause a particular disease. Thus, successful development of diagnostics inevitably depends on the extent of understanding of pathogen diversity. Enteroinvasive *E. coli* infections are now confirmed routinely by genetic analysis[49]. Restricted allelic diversity observed in *Mycobacterium tuberculosis* means that only nominal amino acid variation is expected in proteins of diagnostic or immunoprophylaxis interest. Also, when a new drug is developed it is important to test its efficacy against a set of carefully selected genotypes, which are representative of the population. This necessitates exploration and understanding of the pathogen diversity and this knowledge of pathogen diversity can be exploited in other ways as well. For example, it is known that in-host competition among drug-resistant and drug-sensitive genotypes of a pathogen is pivotal to the spread of drug resistance and hence the clinically useful age of a drug. Simple models show that if drug-sensitive strains are competitively better in host not receiving chemotherapy, the emergence of the drug resistance will be slowed down. However if competitors are removed by chemotherapy, the spread of the resistance is hastened[50]. Such knowledge about the genotypic diversity of a pathogen may be used to assess the effective duration of a newly introduced chemotherapeutic agent against that pathogen.

Vaccine development

Vaccination is the cornerstone of prevention or control of a large number of infectious diseases. Vaccination may also help in eradicating selected infectious diseases. Understanding genetic diversity of microbial pathogens provides a basic framework that can be exploited to rationally design vaccines. Studying comparative genomics of virulent and avirulent strains of a pathogen may help to identify genes encoding putative virulence factors, which should be targeted for successful vaccine development. This is exemplified by *Pasteurella multocida* in which gene for filamentous haemagglutinin was found to be a good candidate for vaccine development by the comparative genomics approach. Similarly, genetic diversity studies successfully argued against the sole use of Osp A protein for the development of vaccine against *Borrelia burgdorferi*, the causative agent of the most common tick-borne infection (Lyme disease) in North America, Europe and Northern Asia[19]. Moreover, the impact of introduction of a new vaccine can be ascertained only when the exact nature of the pathogen population is known. At present, no adequate models are available to predict the likely effects of a vaccination programme that affords protection against only a section of population of a given pathogen. Widespread vaccination with a vaccine against limited range of serogroups of a pathogen can lead to selection of escape variants leading to epidemics. For *N. meningitidis*, mass vaccination with a polyvalent polysaccharide vaccine contributed to the selection of non-serogroup C meningococci leading to outbreaks. Thus, knowledge of genetic diversity is an important pre-requisite for development of successful vaccines.

Long-term benefits of studying pathogen diversity

Understanding evolution and taxonomy

Although first microorganisms appeared on earth more than 3 billion years ago, strictly human pathogens appeared in the last 1.5 million years. Though relatively young, the enormous evolutionary potential of these organisms is illustrated by their innumerable species, ability to infect each and every living host and a variety of life styles. An understanding and knowledge of genetic diversity of pathogens can therefore serve as a rich source for addressing questions related to evolution. Recent studies have shown a parallelism between obligatory pathogens and endosymbionts[51]. Thus, study of the genetic diversity and comparative genomics of microbes associated with human, animal or plant hosts as pathogens or otherwise, may reveal new paradigms of evolution. Genetic diversity studies may allow one to look at the history of evolution or development of a pathogen, and important lessons may be learnt from this for future. For example, in *P. falciparum*, study of genetic variation in mitochondrial DNA supported the contention that major expansion of falciparum malaria occurred about 10,000 years ago. However genetic analysis of nuclear genes did not support this observation[52]. The authors argued that resolving these discrepan-

cies may have important implications for the emergence of drug resistance and vaccine evasion in respect of a new antimalarial drugs or malaria vaccine which may be introduced in future[52].

Nomenclature, classification and taxonomy are most useful only when these are global and reflect hierarchical genetic relationships. These should also be dynamic enough to allow incorporation of new strains, variants and groups[53]. Such robust classification and taxonomy is possible only when data from a large number of genetic variants underlie these developments, and this once again underscores the importance of studying genetic diversity of microorganisms.

Emergence of new pathogens

Genetic diversity studies have provided insights into the origin of several pathogens. An important case in point is that of enterohaemorrhagic *E. coli* (EHEC) or *E. coli* O157 : H7. This important food-borne pathogen has come to notice in the last two decades or so, and has caused large outbreaks of fatal cases of haemolytic uremic syndrome (HUS) and ulcerative colitis. Comparative analysis of 1300 isolates of *E. coli* O157 : H7 and other *E. coli* strains representing 16 serotypes by MLEE, and probing for shiga-like toxin genes revealed that *E. coli* O157 : H7 is closely related to *E. coli* O55 : H7. Lateral acquisition of shiga-toxin and adhesion genes to *E. coli* genome pre-adapted for causing diarrhoea led to the emergence[54,55] of *E. coli* O157 : H7. Similar studies have also indicated that *Mycobacterium tuberculosis* arose ca. 15,000–20,000 years ago from a closely related cattle pathogen *M. bovis* by host specialization. Recent large-scale DNA sequencing results corroborate that *M. tuberculosis* and *M. bovis* have shared a common ancestor in the recent past[56]. Genetic diversity studies also show that *Yersinia pestis*, the causative agent of plague, originated from *Y. pseudotuberculosis* about 1500–20,000 years ago[48].

Bacterial diversity data has also raised certain intriguing questions about the origin of some bacterial pathogens. For example, when mutational frequencies were used to calculate divergence time for *Shigella* from closely related *E. coli*, the estimated time was found to be between 35,000 and 2,70,000 years ago. However, human-specific diseases like shigellosis, which require clusters of human population for transmission and survival, must have originated only after humans developed agriculturally-based large societies, i.e. roughly 10,000 years ago. How *Shigella* would have maintained itself in nature before this period is intriguing and requires further studies[57].

Understanding mechanisms of microbial pathogenicity

The study of genetic diversity of pathogen populations reveals existence of variants having varying degree of pathogenicity. Such studies help in the identification of lineages that have increased propensity to cause disease, viz. identification of hyper-virulent forms of *N. meningitidis* or those of *S. pneumoniae*[13]. This allows one to study the interaction of a pathogen having a spectrum of pathogenicity variants with cells *in vitro* or animal hosts *in vivo*. Such information, in conjunction with the clinical data obtained from patients infected with variant forms of the pathogen is helpful in understanding the exact mechanism of pathogenesis.

Genetic diversity data obtained from the house keeping genes is commonly used to establish phylogenetic relationships among members of a population. This phylogeny can then be used as a framework to which pathogenicity-related characteristics such as virulence, virulence genes and host specificity may be mapped. In this way, important insights can be obtained on the origin of pathogenicity and host adaptations of a pathogen[58]. Moreover, pathogen diversity data when combined with epidemiological, phylogenetic and evolutionary concepts provide insights into the behaviour of pathogens that are unavailable from other simpler approaches[4].

India perspectives

In India, a number of laboratories are involved in the study of the diversity of microbial pathogens prevalent in the region. Pioneering work on the diversity of the Vibrios has been reported from National Institute of Cholera and Enteric Diseases (NICED). *V. cholerae* belonging to serogroups O1 and O139 causes epidemic and pandemic cholera. Ribotyping and pulsed-field gel electrophoresis (PFGE) data showed existence of several clones of *V. cholerae* O1 and O139 serogroups. These clones prevail all over the country and from time to time are replaced by one another. The non-O1, non-O139 serogroups of *V. cholerae*, also known as the non-epidemic serogroups, comprise a heterogeneous group of organisms whose clinical importance is not well understood as yet. Recent studies however clearly show that non-O1, non-O139 serogroups participate in the genesis of new variants of *V. cholerae*. Studies on clonality by RFLP analysis of rRNA genes and PFGE collectively indicate that the recently emerged strains belong to different clones[59]. Considering the importance of these serogroups that can cause diarrhoea by mechanism(s) quite different from that of toxigenic *V. cholerae*, it has been proposed that the nomenclature of enteropathogenic *V. cholerae* should include these serogroups as well. Studies at NICED also showed that despite sharing the same serogroup, environmental and clinical isolates of *V. cholerae* were genetically heterogeneous and were of different lineages[60]. These investigations have also demonstrated the preferential association of heat-stable toxin gene (*stn*) in environmental *V. cholerae* strains belonging to serogroup[61] O14. A comparative study of *Vibrio parahaemolyticus* strains collected from seven countries including India showed uniform pheno-

type and virulence features[62]. Supported by ribotyping and PFGE, it was inferred that the recently emerged O4 : K68 and O1 : KUT strains originated from the pandemic O3 : K 6 clone. A group-specific PCR for the detection of these pandemic strains has also been developed[63]. It is still however not known how these serovar variants of *V. parahaemolyticus* evolved from the initial O3 : K6 pandemic clone and why such an event occurred. These studies have opened up new vistas on the molecular epidemiology of Vibrios.

Genetic diversity of MDR *Mycobacterium tuberculosis* isolated from patients attending various hospitals in northern India was investigated by RFLP associated with IS6110 element[64]. A substantial degree of polymorphism was evident in these strains. Specific regions, *gyr* A and *kat* G were sequenced from a group of the representative isolates to determine their genotypes. Majority were found to belong to Group 1 indicating that evolutionarily these strains were old. Overall, it was concluded that the epidemiological pattern of various strains of *M. tuberculosis* in India was very complex[64]. Sequencing of drug-resistance associated loci in these isolates revealed novel mutations[65] indicating the extent of genetic diversity in these isolates. The Centre for DNA Fingerprinting and Diagnostics (CDFD) has developed a global database of *M. tuberculosis* genotypes based on genome-derived markers. AmpliBASE MT^{TM} is an online databank of high resolution DNA fingerprints representing fluorescent amplified fragment length polymorphism (FAFLP) profiles or amplitypes developed for *M. tuberculosis* complex strains isolated from 48 different countries. AmpliBASE MT^{TM} is based on a SQL database system hyperlinked to visualize genotypic data in the form of DNA fingerprint images of individual strains. A flexible search system based on systematic comparisons of fragment sizes in base pairs allows inter-laboratory comparisons of strain profiles. Besides this, the database also displays previously published data on IS6110 profiles, spoligotypes, MIRU-VNTRs, and large sequence polymorphism along with the FAFLP records to give overall comparisons[66]. Another database called miniBASE MT has been developed for fast strain comparisons based on genomic signatures corresponding to copy numbers at 21 minisatellite like loci (MIRU-VNTR loci) in the *M. tuberculosis* genome. MiniBASE MT has been developed on a MySQL format with a simplified search engine that provides for a faster inter-laboratory exchange of data via Internet. Both the databases are user friendly, freely accessible and are expected to be useful in strengthening the concept of 'geographic genomics' and will be helpful to molecular epidemiologists and those interested in diagnostics development for tuberculosis. These databases can be accessed via http://www.cdfd.org.in. Members of *Mycobacterium avium* complex (MAC) have gained importance due to their association with HIV/AIDS disease. Studies on MAC isolated from India have shown heterogeneity based on plasmid profiling[67].

Genetic diversity of *Helicobacter pylori*, an important pathogen which has been implicated in gastric carcinoma, has been investigated for strains isolated from West Bengal, northern and southern India. A total of 80–90% of the Kolkata strains carried the *cag* pathogenicity island (PAI) and potentially toxigenic *vac As*1 allele, which was independent of the disease status of the patients. This is higher than that in the west where *cag* PAI^{+} *vac As*1 genotypes are disease-associated but lower than that in east Asia[68]. Strains of *H. pylori* isolated from the Santhal and Oroan ethnic minorities of West Bengal, which have long been separated culturally and genetically from mainstream Bengalis, were similar to those isolated from the mainstream Bengalis[69]. Similar results have been reported for North Indian strains too wherein all clinical strains of *H. pylori* carried *vac A* gene. Among these, 69.6% represented genotype s1a, 16.4% genotype s1b and 14% genotype s2. Two different families of *vac A* allele m1 and m2 were also differentiated at the mid-region locus; s1a genotype was present in 86% of the patients with ulcer and 68% of the patients with gastritis[70]. Also, 81–96% of *H. pylori* isolated from north Indian patients possessed *cag A* gene[71]. The heterogeneity of north Indian isolates was also studied by RFLP analysis of *ure C* gene using *Hin*dII, *Alu* I and *Pvu* II, which yielded 15 distinct RFLP patterns[72]. Interestingly, studies on the genetic diversity of *H. pylori* obtained from southern India showed these to be less diverse and these strains were observed to cluster into three groups based on FAFLP[73]. This grouping of *H. pylori* was independent of the disease outcome and the environmental factors such as dietary habits of the patients from which such isolates were obtained. That these groupings may have clonal implications is a novel finding, as opposed to the studies reported for western isolates where *H. pylori* population structure has been reported to be extremely diverse or panmictic[73].

A large number of strains of *Yersinia enterocolitica* isolated from various sources, viz. wastewater, pork, pigs and stools of diarrheic human subject belonged to biotype 1A[74,75]. Though serologically quite diverse, REP- and ERIC-PCR based fingerprinting grouped Indian strains into two clusters which appeared very similar to biotype 1A strains of European and American origin[76]. *Salmonella enterica* serovar paratyphi A of clinical origin showed limited genetic diversity by pulsed-field gel electrophoresis[77,78] and IS200 probing[79]. Amplified ribosomal DNA restriction analysis (ARDRA), RAPD and BOX-PCR have also been used for typing Indian strains of *Salmonella*[80]. Uniform ARDRA patterns were reported for all strains studied including serotypes typhi, paratyphi A and typhimurium. Although BOX-PCR differentiated paratyphi A from the others, RAPD was most discriminatory for typing *Salmonella* strains. The genetic diversity of 42 multiple drug-resistant isolates of *Pseudomonas aeruginosa* from 11 patients with post-surgical endopthalmitis was studied using FAFLP fingerprinting[81]. Isolation of more

than one type of strains from each of the patients suggested mixed infection, with the possibility of cross-infection by at least three different strains. Leptospirosis is a major public health problem in Andaman Islands. *Leptospira* isolates recovered recently from human cases were compared with one of the earliest available isolates from these islands, that date back to 1929. RAPD analysis indicated that some of the recent isolates (2001 AD) were identical to those isolated in 1929, suggesting long-term persistence of the genotypes of *Leptospira* involved in the causation of acute leptospirosis of humans in these islands[82].

Genetic variability of plant pathogen *Xanthomonas oryzae* pv. *oryzae* isolates was studied using RAPD and IS1112-based PCR which revealed heterogeneity in the isolates collected from different parts of the country[83].

Studies on the genetic diversity of certain viruses causing important human diseases such as HIV-1, hepatitis C virus (HCV) and rotaviruses have also been reported from India. Surveillance studies[84] on HIV-1 showed that 78.4% isolates belonged to subtype C, 8.8% to subtype B, 2.4% to subtype A and 1.6% to subtype E. In this study HIV-1 subtypes were determined for homologies in the C2-V3-V5 region by heteroduplex mobility assay (HMA). In some isolates, subtype determination was not unequivocal indicating possibile existence of recombinant strains of HIV-1. Extensive studies on the genotypic diversity of Group A rotaviruses isolated from diarrheic human subjects from different parts of India have been conducted. These studies reveal that predominant strains[85] are G1 to G4 except those from northern and central India where G9 is the predominant G type[86]. Apart from the overall diversity of rotaviruses, a marked difference in the geographic distribution of strains within India was also observed. In some Indian cities, as many as eight strains were circulating at the same time, a diversity far greater than that seen in most other countries[87]. Three new human G12 strains have recently been detected from diarrhoeic clinical samples of children in Kolkata during a routine surveillance study of rotaviral diarrhoea in India[88]. Such genetic diversity studies on rotaviruses in India have clear implications for vaccine development. Genotypic determination of hepatitis C virus (HCV) isolated from northern and southern India has also been reported. Eleven northern Indian isolates were genotyped using sequence comparison of part of the non-structural (NS 5) and structural (core) regions. Four isolates showed sequence homology to type 1b. Two of the isolates were classified as type 3a. One isolate was classified as type 3b and the remaining four isolates were found to be variants of type 3 but did not belong to any designated subtype. Two of the unclassified isolates were put into a new subtype of 3, named 3g. This study demonstrated that type 3 variants including a new subtype (3g) of HCV were predominant in northern India[89]. Contrary to this, of the 24 isolates from southern India, 21 corresponded to HCV type 1 and others to type 3. Evidence was also presented to show that two of the type 3 isolates may represent novel subtypes. These data, in general, indicated that HCV genotype 1 predominated over HCV genotype 3 in southern India[90]. A study on Norwalk-like viruses (NLVs) isolated from a food-borne outbreak of acute gastroenteritis in Delhi placed the viruses in Toronto virus cluster of GG II NLVs based on sequencing of RdRp gene[91].

Plant viruses, which have been investigated from the point of view of genetic diversity, include yellow mosaic virus infecting soybean[92] and cotton leaf curl virus[93]. Limited studies have been reported from India on the genetic diversity of eukaryotic pathogenic microorganisms – whether protozoan parasites or fungi. Studies carried out with *Entamoeba histolytica*, showed that probes derived from ITS 1 and ITS 2 are more reliable in distinguishing[94] it from the non-pathogen *E. dispar*. Another sequence identified during expressed sequence tag analysis of *E. dispar* showed clear divergence in sequence at some regions with the *E. histolytica* homologues[95], which may be used to differentiate the two species. The diversity of *E. histolytica*[96] and *Giardia lamblia*[97] isolates from Delhi has been investigated by zymodeme analysis. Genetic relatedness of twelve clinical and five environmental isolates of *Cryptococcus neoformans* var. *neoformans* isolated from Tamil Nadu was studied by RAPD which clustered these isolates into three groups[98].

The rice blast fungus *Magnaporthe grisea* isolates collected from southern Karnataka clustered into three lineages based on RAPD analysis[99]. *Colletotrichum graminicola* obtained from six provinces of India was analysed by PCR-amplification of intergenic spacer region of nuclear rDNA. No definite relationships could be discerned between the genotype and the source/location of isolation[100]. Diversity of *Phytophthora palmivora* from cocoa and coconut, and *P. capsici* from black pepper was studied by AFLP analysis which showed similar pattern for isolates from two plants. Among the isolates of *P. capsici*, four AFLP fingerprint groups were evident which were distinct from that of *P. palmivora*[101]. Limited studies have also been reported with *P. nicotianae* causing leaf rot of betelvine[102] and *Trichoderma* isolates obtained from two Indian type culture collection centers[103]. This shows that a number of laboratories from India have contributed to the study of genetic diversity of pathogens extant in this region of the globe. However to quote Mark Achtman[53] – 'A considerable portion of the global diversity of bacterial pathogens is quite likely to be present in developing countries that have not yet been sampled extensively. The epidemiology of bacterial diseases in the developed world is not representative of the whole globe….'

In conclusion, we are just beginning to understand the genetic diversity and population genetics of pathogenic microorganisms. Such studies should preferably be undertaken with structured collections of the pathogens using highly portable, state-of-the-art technique such as multilocus sequence typing which is also amenable to validation

and electronic data-basing so that global comparisons can be made. However, using data already available, considerable insights have already been gained into the molecular mechanisms underlying the variations observed in the frequency and severity of infectious diseases, disparate behaviour of the clonal lineages of pathogens, and the emergence of new and drug-resistant microbes. Such studies have also the potential to address problems relevant to public health such as understanding nosocomial outbreaks, global epidemiology, diagnostics, and drugs and vaccine development. It is expected that improved understanding of the pathogen diversity will enable us to practice public health interventions with greater sophistication. Long-term benefits, which may accrue from such studies, include clues to the emergence of new pathogens and improved understanding of microbial pathogenicity *per se*, and evolution and taxonomy. A large number of methods currently available are being used to index genome-wide variations in pathogenic microbes especially prokaryotic. However, peculiarities inherent in the genetic diversity of eukaryotic pathogens such as pathogenic fungi and parasitic protozoa are still poorly understood. The complete genome sequences of a number of pathogens have already become available and more are being sequenced. With high throughput DNA sequencing becoming cheaper and easily available, genetic diversity studies will get further impetus and would be one of our major allies in fight against pathogenic microorganisms.

1. Delong, E. F., Marine microbial diversity: The tip of the iceberg. *Trends Biotechnol*., 1997, **15**, 203–207.
2. Hugenholtz, P., Goebel, B. M. and Pace, N. R., Impact of culture-independent studies on the emerging phylogenetic view of bacterial diversity. *J. Bacteriol.*, 1998, **180**, 4765–4774.
3. Kroes, I., Lepp, P. W. and Relman, D. A., Bacterial diversity within the human subgingival crevice. *Proc. Natl. Acad. Sci. USA*, 1999, **96**, 14547–14552.
4. Maiden, M. C. J., High-throughput sequencing in the population analysis of bacterial pathogens of humans. *Int. J. Med. Microbiol.*, 2000, **290**, 183–190.
5. Gupta, S. and Maiden, M. C. J., Exploring the evolution of diversity in pathogen populations. *Trends Microbiol*., 2001, **9**, 181–185.
6. Singleton, P., *DNA Methods in Clinical Microbiology*, Kluwer, Dordrecht, The Netherlands, 2000.
7. Wong, C. W. *et al.*, Tracking the evolution of the SARS coronavirus using high throughput high-density resequencing arrays. *Genome Res*., 2004, **14**, 398–405.
8. Selander, R. K., Caugant, D. A., Ochman, H., Musser, J. M., Gilmour, M. N. and Whittam, T. S., Methods of multilocus enzyme electrophoresis for bacterial population genetics and systematics. *Appl. Environ. Microbiol*., 1986, **51**, 873–884.
9. Kimura, M., *The Neutral Theory of Molecular Evolution*, Cambridge University Press, London, 1983.
10. Hartl, D. L., Medhora, M., Green, L. and Dykhuizen, D. E., The evolution of DNA sequences in *Escherichia coli*. *Philos. Trans. R. Soc. London B Biol. Sci.*, 1986, **312**, 191–204.
11. Milkman, R. and Bridges, M. M., Molecular evolution of the *Escherichia coli* chromosome. III. clonal frames. *Genetics*, 1990, **126**, 505–517.
12. Nelson, K., Schlievert, P. M., Selander, R. K. and Musser, J. M., Characterization and clonal distribution of four alleles of the *speA* gene encoding pyrogenic exotoxin A (scarlet fever toxin) in *Streptococcus pyrogenes*. *J. Exp. Med*., 1991, **174**, 1271–1274.
13. Maiden, M. C. J. *et al*., Multilocus sequence typing: a portable approach to the identification of clones within populations of pathogenic microorganisms. *Proc. Natl. Acad. Sci. USA*, 1998, **95**, 3140–3145.
14. Orskov, F. and Orskov, I., Summary of a workshop on the clone concept in the epidemiology, taxonomy and evolution of the *Enterobacteriaceae* and other bacteria. *J. Infect. Dis*., 1983, **148,** 346–357.
15. Smith, J. M., Smith, N. H., O'Rourke, M. and Spratt, B. G., How clonal are bacteria? *Proc. Natl. Acad. Sci. USA*, 1993, **90**, 4384–4388.
16. Spratt, B. G., Smith, N. H., Zhou, J., O'Rourke, M. and Feil, E., The population genetics of the pathogenic *Neisseria*. In *Population Genetics of Bacteria* (eds Baumberg, S., Young, J. P. W., Wellington, E. M. H. and Saunders, J. R.), Cambridge University Press, London, 1995, pp. 143–160.
17. Holmes, E. C., Urwin, R. and Maiden, M. C. J., The influence of recombination on the population structure and evolution of human pathogen *Neisseria meningitidis*. *Mol. Biol. Evol*., 1999, **16**, 741–749.
18. Boyd, E. F., Wang F. S., Whittam, T. S. and Selander, R. K., Molecular genetic relationships of the salmonellae. *Appl. Environ. Microbiol*., 1996, **62**, 804–808.
19. Dykhuizen, D. E., Polin, D. S., Dunn, J. J., Wilske, B., Preac–Mursic, V., Dattwyler R. J. and Luft, B.J., *Borrelia burgdorferi* is clonal: implications for taxonomy and vaccine development. *Proc. Natl. Acad. Sci. USA*, 1993, **90**, 10163–10167.
20. Tibayrenc, M., Towards a unified evolutionary genetics of microorganisms. *Annu. Rev. Microbiol*., 1996, **50**, 401–429.
21. Tibayrenc, M., Kjellberg, F. and Ayala, F. J., A clonal theory of parasitic protozoa: the population structures of *Entamoeba, Giardia, Leishmania, Naegleria, Plasmodium, Trichomonas*, and *Trypanosoma* and their medical and taxonomical consequences. *Proc. Natl. Acad. Sci. USA*, 1990, **87**, 2414–2418.
22. Day, K. P., Koella, J. C., Nee, S., Gupta, S. and Read, A. F., Population genetics and dynamics of *Plasmodium falciparum*: an ecological view. *Parasitology*, 1992, **104**, Suppl, S35–S52.
23. Mazars, E. and Dei-Cas, E., Epidemiological and taxonomic impact of *Pneumocystis* biodiversity. *FEMS Immunol. Med. Microbiol.*, 1998, **22**, 75–80.
24. Sokurenko, E. V. *et al*., Pathogenic adaptation of *Escherichia coli* by natural variation of the Fim H adhesin. *Proc. Natl. Acad. Sci. USA*, 1998, **95**, 8922–8926.
25. Boucher, J. C., Yu, H., Mudd, M. H. and Deretic, V., Mucoid *Pseudomonas aeruginosa* in cystic fibrosis: characterization of *muc* mutations in clinical isolates and analysis of clearance in a mouse model of respiratory infection. *Infect. Immun.*, 1997, **65**, 3838–3846.
26. Yu, H., Hanes, M., Chrisp, C. E., Boucher, J. C. and Deretic, V., Microbial pathogenesis in cystic fibrosis: pulmonary clearance of mucoid *Pseudomonas aeruginosa* and inflammation in a mouse model of repeated respiratory challenge. *Infect. Immun.*, 1998, **66**, 280–288.
27. Maurelli, A. T., Fernandez, R. E., Bloch, C. A., Rode, C. K. and Fasano, A., Black holes' and bacterial pathogenicity: a large genomic deletion that enhances the virulence of *Shigella* spp. and enteroinvasive *Escherichia coli. Proc. Natl. Acad. Sci. USA*, 1998, **95**, 3943–3948.
28. Berg, D. E., Transposon Tn5. In *Mobile DNA* (eds Berg, D. E. and Howe, M. M.), American Society for Microbiology, Washington DC, 1989, pp. 185–210.
29. Hall, R. M., Mobile gene cassettes and integrons: moving antibiotic resistance genes in Gram-negative bacteria. *CIBA Foundation Symp*., 1997, **207**, 192–205.
30. Rowe-Magnus, D. A. and Mazel, D., Resistance gene capture. *Curr. Opin. Microbiol*., 1999, **2**, 483–488.

31. Mazel, D., Dychinco, B., Webb, V. A. and Davies, J. A., Distinctive class of integron in the *Vibrio cholerae* genome. *Science*, 1998, **280**, 605–608.
32. Feil, E. J. *et al*., Recombination within natural populations of pathogenic bacteria: short-term empirical estimates and long-term phylogenetic consequences. *Proc. Natl. Acad. Sci. USA*, 2001, **98**, 182–187.
33. Ochman, H., Lawrence, J. G. and Groisman, E. A., Lateral gene transfer and the nature of bacterial innovation. *Nature*, 2000, **405**, 299–304.
34. Dowson, C. G. *et al*., Horizontal transfer of penicillin-binding protein genes in penicillin-resistant clinical isolates of *Streptococcus pneumoniae*. *Proc. Natl. Acad. Sci. USA*, 1989, **86**, 8842–8846.
35. Dowson, C. G., Hutchison, A., Woodford, N., Johnson, A. P., George, R. C. and Spratt, B. G., Penicillin-resistant viridans streptococci have obtained altered penicillin-binding protein genes from penicillin-resistant strains of *Streptococcus pneumoniae*. *Proc. Natl. Acad. Sci. USA*, 1990, **87**, 5858–5862.
36. Coffey, T. J., Dowson, C. G., Daniels, M., Zhou, J., Martin, C., Spratt, B. G. and Musser, J. M., Horizontal transfer of multiple penicillin-binding protein genes, and capsular biosynthetic genes, in natural populations of *Streptococcus pneumoniae*. *Mol. Microbiol*., 1991, **5**, 2255–2260.
37. Lomholt, H., Evidence of recombination and an antigenically diverse immunoglobulin A1 protease among strains of *Streptococcus pneumoniae*. *Infect. Immun.*, 1995, **63**, 4238–4243.
38. Musser, J. M. and Kapur, V., Clonal analysis of methicillin resistant *Staphylococcus aureus* strains from intercontinental sources: association of the *mec* gene with divergent phylogenetic lineages implies dissemination by horizontal transfer and recombination. *J. Clin. Microbiol*., 1992, **30**, 2058–2063.
39. Boer, P., Wagenaar, J. A., Achterberg, R. P., Putten, J. P., Schouls, L. M. and Duim, B., Generation of *Campylobacter jejuni* genetic diversity *in vivo*. *Mol. Microbiol*., 2002, **44**, 351–359.
40. Waldor, M. K. and Mekalanos, J. J., Lysogenic conversion by a filamentous phage encoding cholera toxin. *Science*, 1996, **272**, 1910–1914.
41. Lindsay, J. A., Ruzin, A., Ross, H. F., Kurepina, N. and Novick, R. P., The gene for toxic shock toxin is carried by a family of mobile pathogenicity islands in *Staphylococcus aureus*. *Mol. Microbiol*., 1998, **29**, 527–543.
42. Mesyanzhinov, V. V. *et al*., The genome of bacteriophage *f*KZ of *Pseudomonas aeruginosa*. *J. Mol. Biol*., 2002, **317**, 1–19.
43. Hacker, J. and Kaper, J. B., Pathogenicity islands and the evolution of microbes. *Annu. Rev. Microbiol*., 2000, **54**, 641–679.
44. Hacker, J. and Carniel, E., Ecological fitness, genomic islands and bacterial pathogenicity. A Darwinian view of the evolution of microbes. *EMBO Rep*., 2001, **2**, 376–381.
45. Hacker, J., Blum-Oehler, G., Muhldorfer, I. and Tschape, H., Pathogenicity islands of virulent bacteria: structure, function and impact on microbial evolution. *Mol. Microbiol*., 1997, **23**, 1089–1097.
46. Buchrieser, C., Brosch, R., Bach, S., Guiyoule, A. and Carniel, E., The high-pathogenicity island of *Yersinia pseudotuberculosis* can be inserted into any of the three chromosomal *asn* tRNA genes. *Mol Microbiol*., 1998, **30**, 965–978.
47. Janke, B., Dobrindt, U., Hacker, J. and Blum-Oehler, G., A subtractive hybridisation analysis of genomic differences between the uropathogenic *E. coli* strain 536 and the *E. coli* K-12 strain MG1655. *FEMS Microbiol. Lett*., 2001, **199**, 61–66.
48. Achtman, M., Zurth, K., Morelli. G., Torrea. G., Guiyoule, A. and Carniel, E., *Yersinia pestis*, the cause of plague, is a recently emerged clone of *Yersinia pseudotuberculosis*. *Proc. Natl. Acad. Sci. USA*, 1999, **96**, 14043–14048.
49. Bopp, C. A., Brenner, F., Wells, J. G. and Strockbine, N. A., *Escherichia, Shigella* and *Salmonella.* In *Manual of Clinical Microbiology* (eds Murray, P. R., Baron, E. J., Pfaller, M. A., Tenover, F. C. and Yolken, R. H.), ASM Press, Washington DC, 1999, 7th edn, pp. 459–474.
50. Read, A. F. and Taylor, L. H., The ecology of genetically diverse infections. *Science*, 2001, **292**, 1099–1102.
51. Ochman, H. and Moran, N. A., Genes lost and genes found: evolution of bacterial pathogenesis and symbiosis. *Science*, 2001, **292**, 1096–1099.
52. Hartl, D. L., The origin of malaria: mixed messages from genetic diversity. *Nature Rev. Microbiol*., 2004, **2**, 15–22.
53. Achtman, M., A phylogenetic perspective on molecular epidemiology. In *Molecular Medical Microbiology* (ed. Sussman, M.), Academic Press, New York, 2001, pp. 1–25.
54. Whittam, T. S., Wolfe, M. L., Wachsmuth, I. K., Orskov, F., Orskov, I. and Wilson, R. A., Clonal relationships among *Escherichia coli* strains that cause hemorrhagic colitis and infantile diarrhea. *Infect. Immun.*, 1993, **61**, 1619–1629.
55. Whittam, T. S., Genetic population structure and pathogenicity in enteric bacteria. In *Population Genetics of Bacteria* (eds Baumberg, S., Young, J. P. W., Saunders, S. R. and Wellington, E. M. H.), Cambridge University Press, London, 1995, pp. 217–245.
56. Musser, J. M., Molecular population genetic analysis of emerged bacterial pathogens: selected insights. *Emerg. Infect. Dis.*, 1996, **2**, 1–17.
57. Stephens, C. and Murray, W., Pathogen evolution: how good bacteria go bad. *Curr. Biol*., 1999, **11**, R53–R56.
58. Spratt, B. G. and Maiden M. C. J., Bacterial population genetics, evolution and epidemiology. *Philos. Trans. Soc. London B Biol. Sci.*, 1999, **354**, 701–710.
59. Sharma, C. *et al.*, Molecular analysis of non-O1, non-O139 *Vibrio cholerae* associated with an unusual upsurge in the incidence of cholera-like disease in Calcutta, India. *J. Clin. Microbiol.*, 1998, **36**, 756–763.
60. Chakraborty, S. *et al*., Comparison of antibiogram, virulence genes, ribotypes and DNA fingerprints of *Vibrio cholerae* of matching serogroups isolated from hospitalised diarrhoea cases and from the environment during 1997–1998 in Calcutta, India. *J. Med. Microbiol*., 2001, **50**, 879–888.
61. Sarkar, B., Bhattacharya, T., Ramamurthy, T., Shimada, T., Takeda Y. and Nair, G. B., Preferential association of the heat-stable enterotoxin gene (*stn*) with environmental strains of *Vibrio cholerae* belonging to the O14 serogroup. *Epidemiol. Infect.*, 2002, **129**, 245–251.
62. Chowdhury, N. R., Chakraborty, S., Ramamurthy, T., Nishibuchi, M., Yamasaki, S., Takeda, Y. and Nair, G. B., Molecular evidence of clonal *Vibrio parahaemolyticus* pandemic strains. *Emerg. Infect. Dis*., 2000, **6**, 631–636.
63. Matsumoto, C. *et al*., Pandemic spread of an O3:K6 clone of *Vibrio parahaemolyticus* and emergence of related strains evidenced by arbitrarily primed PCR and toxRS sequence analyses. *J. Clin. Microbiol*., 2000, **38**, 578–585.
64. Siddiqi, N. *et al.*, Typing of drug resistant isolates of *Mycobacterium tuberculosis* from India using the IS6110 element reveals substantive polymorphism. *Infect. Genet. Evol*., 2001, **1**, 109–116.
65. Siddiqi, N. *et al.*, Molecular characterization of multidrug-resistant isolates of *Mycobacterium tuberculosis* from patients in north India. *Antimicrob. Agents Chemother*., 2002, **46**, 443–450.
66. Majeed, A. A. *et al.*, AmpliBASE MT^{TM}: a *Mycobacterium tuberculosis* diversity knowledgebase. *Bioinformatics*, 2004, **20**, 989–992.
67. Dam, T., Bose, M., Isa, M. and Virdi, J. S., Isolation of plasmids from *Mycobacterium avium – intracellulare Complex* (MAC) strains from India. *J. Med. Microbiol.*, 2000, **49**, 392–393.
68. Mukhopadhyay, A. K. *et al*., Distinctiveness of genotypes of *Helicobacter pylori* in Calcutta, India. *J. Bacteriol*., 2000, **182**, 3219–3227.
69. Datta, S. *et al.*, Virulence genes and neutral DNA markers of *Helicobacter pylori* isolates from different ethnic communities of West Bengal, India. *J. Clin. Microbiol*., 2003, **41**, 3737–3743.

70. Mishra, K. K., Srivastava, S., Dwivedi, P. P., Prasad, K. N. and Ayyagari, A., Genotypes of *Helicobacter pylori* isolated from various acid peptic diseases in and around Lucknow. *Curr. Sci*., 2002, **83**, 749–755.
71. Mishra, K. K., Srivastava, S., Dwivedi, P. P., Prasad, K. N. and Ayyagari, A., *ure* C PCR based diagnosis of *Helicobacter pylori* infection and detection of *cag* A gene in gastric biopsies. *Indian J. Pathol. Microbiol.*, 2002, **45**, 31–37.
72. Mishra, K. K., Srivastava, S., Dwivedi, P. P., Prasad, K. N. and Ayyagari, A., PCR-based RFLP analysis of *ure* C gene for typing of Indian *Helicobacter pylori* strains from gastric biopsy specimens and culture. *J. Microbiol.*, 2002, **40**, 282–288.
73. Ahmed, N. *et al.*, Genomic analysis of *Helicobacter pylori* from Andhra Pradesh, South India: Molecular evidence for three major genetic clusters. *Curr. Sci.*, 2003, **85**, 1579–1586.
74. Sinha, I., Choudhary, I. and Virdi, J. S., Isolation of *Yersinia enterocolitica* and *Yersinia intermedia* from wastewater and their biochemical and serological characteristics. *Curr. Sci*., 2000, **79**, 510–513.
75. Singh, I., Bhatnagar, S. and Virdi, J. S., Isolation and characterization of *Yersinia enterocolitica* from diarrheic human subjects and other sources. *Curr. Sci*., 2003, **84**, 1353–1355.
76. Sachdeva, P. and Virdi, J. S., Repetitive elements sequence (REP/ERIC)-PCR based genotyping of clinical and environmental strains of *Yersinia enterocolitica* biotype 1A reveal existence of limited number of clonal groups. *FEMS Microbiol. Lett.*, 2004, **240**, 193–201.
77. Thong, K. L. *et al.*, Molecular analysis of *Salmonella paratyphi* A from an outbreak in New Delhi, India. *Emerg. Infect. Dis*., 1998, **4**, 507–508.
78. Goh, Y. L. *et al.*, Genetic diversity of *Salmonella enterica* serovar paratyphi A from different geographical regions of Asia. *J. Appl. Microbiol*., 2002, **92**, 1167–1171.
79. Chandel, D. S., Nisar, N., Thong, K. L. Pang, T. and Chaudhry, R., Role of molecular typing in an outbreak of *Salmonella paratyphi* A. *Trop. Gastroenterol*., 2000, **21**, 121–123.
80. Tikoo, A., Tripathi, A. K., Verma, S. C., Agrawal, N. and Nath, G., Application of PCR fingerprinting techniques for identification and discrimination of *Salmonella* isolates. *Curr. Sci*., 2001, **80**, 1049–1052.
81. Ahmed, N. *et al.*, Whole genome fingerprinting and genotyping of multiple drug resistant (MDR) isolates of *Pseudomonas aeruginosa* from endophthalmitis patients in India. *Infect. Genet. Evol*., 2002, **1**, 237–242.
82. Roy, S., Biswas, D., Vijaychari, P., Sugunan, A.P. and Sehgal, S. C., Antigenic and genetic relatedness of *Leptospira* strains isolated from the Andaman islands in 1929 and 2001. *J. Med. Microbiol*., 2003, **52**, 909–911.
83. Gupta, V. S., Rajebhosale, M. D., Sodhi, M., Singh, S., Gnanamanickam, S. S., Dhaliwal, H. S. and Ranjekar, P. K., Assessment of genetic variability and strain identification of *Xanthomonas oryzae* pv. *oryzae* using RAPD–PCR and IS1112-based PCR. *Curr. Sci*., 2001, **80**, 1043–1048.
84. Sahni, A. K., Prasad, V. V. and Seth, P., Genomic diversity of human immunodeficiency virus type-1 in India. *Int. J. STD AIDS*, 2002, **13**, 115–118.
85. Das, S. *et al.*, Emergence of novel human group A rotavirus G12 strains in India. *J. Clin. Microbiol*., 2003, **41**, 2760–2762.
86. Ramachandran, M. *et al*., Unusual diversity of human rotavirus G and P genotypes in India. *J. Clin. Microbiol.*, 1996, **34**, 436–439.
87. Jain, V., Das, B. K., Bhan, M. K., Glass, R. I. and Gentsch, J. R., The Indian strain surveillance collaborating laboratories. Great diversity of group A rotavirus strains and high prevalence of mixed rotavirus infections in India. *J. Clin. Microbiol*., 2001, **39**, 3524–3529.
88. Das, S. *et al.*, Genomic diversity of group A rotavirus strains infecting humans in eastern India. *J. Clin. Microbiol*., 2002, **40**, 146–149.
89. Panigrahi, A. K., Roca, J., Acharya, S. K., Jameel, S. and Panda, S. K., Genotype determination of hepatitis C virus from northern India: identification of a new subtype. *J. Med. Virol*., 1996, **48**, 191–198.
90. Valliammai, T., Thyagarajan, S. P., Zuckerman, A. J. and Harrison, T. J., Diversity of genotypes of hepatitis C virus in southern India. *J. Gen. Virol.*, 1995, **76**, 711–716.
91. Girish, R., Broor, S., Dar, L. and Ghosh, D., Foodborne outbreak caused by a Norwalk-like virus in India. *J. Med. Virol*., 2002, **67**, 603–607.
92. Usharani, K. S., Surendranath, B., Haq, Q. M. R. and Malathi, V. G., Yellow mosaic virus infecting soybean in northern India is distinct from the species infecting soybean in southern and western India. *Curr. Sci*., 2004, **86**, 845–850.
93. Narayananaswamy, K., Hemalatha, S. P., Priyadarshini, P., Rishi, N. and Savithri, H. S., Molecular characterization of cotton leaf curl virus isolates from India. 10th Congress of Federation of Asian and Oceanian Biochemists and Molecular Biologists, Bangalore, India, December 7–11, 2003.
94. Paul, J., Bhattacharya, A. and Bhattacharya, S., Close sequence identity between ribosomal DNA episomes of the non-pathogenic *Entamoeba dispar* and pathogenic *Entamoeba histolytica*. *J. Biosci.*, 2002, **6**, 619–627.
95. Azam, A., Paul, J., Sehgal, D., Prasad, J., Bhattacharya, S. and Bhattacharya, A., Identification of novel genes from *Entamoeba histolytica* by expressed sequence tag analysis. *Gene*, 1996, **181**, 113–116.
96. Baveja, U. K., Francis, S., Kaur, M. and Agarwal, S. K., The zymodemes of *Entamoeba histolytica* in New Delhi, India. *J. Diarrhoel Dis. Res*., 1990, **8**, 27–30.
97. Baveja, U. K., Jyoti, A. S., Kaur, M., Agarwal, D. S., Anand, B. S. and Nanda, R., Isoenzyme studies of *Giardia lamblia* isolated from symptomatic cases. *Aust. J. Exp. Biol. Med. Sci*., 1986, **64**, 119–126.
98. Kidd, S. E. *et al.*, Molecular epidemiology of clinical and environmental isolates of *Cryptococcus neoformans* from Tamil Nadu, India. *Austr. Mycol.*, 2001, **20**, 105–114.
99. Srinivasachary, Hittalmani, S., Shivayogi, S., Vaishali, M. G., Shashidhar, H. E. and Kumar, K. G., Genetic analysis of rice blast fungus of southern Karnataka using DNA markers and reaction of popular rice genotypes. *Curr. Sci.*, 2002, **82**, 732–735.
100. Latha, J., Chakrabarti, A., Mathur, A., Rao, V. P., Thakur, R. P. and Mukherjee, P. K., Genetic diversity of *Colletotrichum graminicola* isolates from India revealed by restriction analysis of PCR-amplified intergenic spacer region of nuclear rDNA. *Curr. Sci.*, 2003, **84**, 881–883.
101. Chowdappa, P., Brayford, D., Smith, J. and Flood, J., Molecular discrimination of *Phytophthora* isolates on cocoa and their relationship with coconut, black pepper and bell pepper isolates based on rDNA repeat and AFLP fingerprints. *Curr. Sci*., 2003, **84**, 1235–1238.
102. Tripathi, A., Singh, R., Raj, S. K., Singh, A. P. and Johri, J. K., Molecular identification of *Phytophthora nicotianae* isolates causing leaf rot of betalvine (*Piper betel* L.). *Curr. Sci*., 2003, **84**, 22–24.
103. Latha, J., Verma, A. and Mukherjee, P. K., PCR-fingerprinting of some *Trichoderma* isolates from two Indian type culture collections – a need for re-identification of these economically important fungi. *Curr. Sci*., 2002, **83**, 372–374.

ACKNOWLEDGEMENTS. JSV thanks Drs Niyaz Ahmed (CDFD, Hyderabad), Archana Ayyagari (SGPGIMS, Lucknow), Usha K. Baveja (NICD, Delhi), Mridula Bose (VPCI, Delhi), Shobha Broor (AIIMS, Delhi), Rama Chaudhry (AIIMS, Delhi), T. N. Naik (NICED, Kolkata), Jaishree Paul (JNU, New Delhi), Gaya Prasad (CCSHAU, Hisar), T. Ramamurthy (NICED, Kolkatta), S. P. Thyagarajan (Dr. ALMPGIBMS, Chennai) and S. Vijayachari (RMC, Port Blair) for inputs. Work on microbial pathogens in author's laboratory is supported by financial assistance from DST, DBT, DRDO, MoEF and ICMR.