Research Article

Multi locus sequence typing of *Burkholderia pseudomallei* isolates from India unveils molecular diversity and confers regional association in Southeast Asia

Veeraraghavan Balaji1*, Susmitha Perumalla1, Rajamani Perumal2, Francis Yesurajan Inbanathan1, Suresh Kumar Rajamani Sekar1, Miracle Magdelene Paul1, Rani Diana Sahni1, John Antony Jude Prakash1, Ramya Iyadurai1

1 Department of Clinical Microbiology, Christian Medical College, Vellore, Tamilnadu, India, 2 Department of Orthopaedics, Christian Medical College, Vellore, Tamilnadu, India, 3 Department of Medicine, Christian Medical College, Vellore, Tamilnadu, India

* vbalaji@cmcvellore.ac.in

Abstract

Objectives

*Burkholderia pseudomallei*, the causative agent for melioidosis, has become a public health problem in India and across the world. Melioidosis can be difficult to diagnose because of the inconsistent clinical presentations of the disease. This study aims to determine the genetic diversity among the clinical isolates of *B. pseudomallei* from India in order to establish a molecular epidemiology and elucidate the Southeast Asian association.

Methods

Molecular typing using multi locus sequence typing was performed on thirty one archived *B. pseudomallei* clinical isolates, previously characterised from specimens obtained from patients admitted to the Christian Medical College & Hospital, Vellore from 2015 to 2016. Further investigations into the genetic heterogeneity and evolution at a regional and global level were performed using *insilico* tools.

Results

Multi locus sequence typing (MLST) of the isolates from systemic and localized forms of melioidosis, including blood, pus, tissue, and urine specimens, revealed twenty isolates with novel sequence types and eleven with previously reported sequence types. High genetic diversity was observed using MLST with a strong association within the Southeast Asian region.

Conclusions

Molecular typing of *B. pseudomallei* clinical isolates using MLST revealed high genetic diversity and provided a baseline molecular epidemiology of the disease in India with a
strong Southeast Asian association of the strains. Future studies should focus on whole genome based Single-Nucleotide-Polymorphism (SNP) which has the advantage of a high discriminatory power, to further understand the novel sequence types reported in this study.

Author summary

Burkholderia pseudomallei, a gram negative bacteria, is the causative agent for melioidosis. Annually, around 165,000 people suffer from melioidosis worldwide. B. pseudomallei is present in wet soil and stagnant water. It enters the human body via percutaneous inoculation, inhalation, aspiration, and occasionally ingestion. Clinical presentations of B. pseudomallei vary by geographical region. Melioidosis occurs predominantly in Southeast Asia, northern Australia, South Asia (including India), and China. Occasional cases occur in other countries around the world. Melioidosis has become a public health problem in India, due to the increasing numbers of people affected in various parts of the country. This study provides baseline data on the genetic diversity among B. pseudomallei isolates from different clinical samples (blood, pus, tissue and urine) of patients admitted to a tertiary care hospital using signature nucleotide sequences via multi locus sequence typing (MLST). Further, this study shows a relationship among B. pseudomallei previously reported in various Southeast Asian countries over the years from 1935 and 1947 with those seen in current clinical cases.

Introduction

Burkholderia pseudomallei, the causative agent of the infectious disease melioidosis, is estimated to cause 165,000 cases of human melioidosis per year worldwide [1]. B. pseudomallei, an environmental saprophyte is commonly found in wet soil and stagnant water throughout endemic regions. The mode of infection is by inhalation, through cuts in the skin, and occasionally through ingestion [2]. The most severe clinical manifestation is melioidosis septic shock, which is often associated with pneumonia and bacterial dissemination to distant sites [3]. Melioidosis often affects individuals with one or more pre-existing conditions associated with an altered immune response, the most common being diabetes mellitus [4,5].

Melioidosis is endemic to Southeast Asia and Northern Australia, but still can be recognized in other countries worldwide [6]. Regional variations in clinical presentation of melioidosis are widely observed such as the predominance of pivotal swelling seen in Australia but not in Thailand. The contributing factors for this diversity are still unclear whether bacterial, host or environmental. There is no correlation between the clinical presentations and genotypes to date, even though environmental partitioning between Australian and Asian population of B. pseudomallei have been reported previously [7]. Melioidosis has become a public health problem in India, due to the steady rise in case detection rates from various parts of the country. Moreover, no consistency has been observed in the forms of melioidosis (clinical presentations) reported among the sporadic cases across the country in the last two decades. A recent report reveals genetic diversity among clinical isolates of B. pseudomallei from South India [8]. This study focuses on the clinical manifestations and genetic diversity of B. pseudomallei isolated from patients across India using the multi locus sequence typing (MLST) scheme for B. pseudomallei as described by PubMLST, with an attempt to establish the molecular epidemiology in Southeast Asian region.
Methods

A total of 31 *B. pseudomallei* clinical isolates that were previously characterised from different clinical specimens (blood, pus, tissue and urine) obtained from patients admitted to the Christian Medical College & Hospital, Vellore from different parts of the country, during 2015 to 2016 were included in this study. The total genomic DNA was extracted using automated method (QIASeqSymphony SP, QIAGEN, Germany). MLST was performed by PCR amplification of seven house-keeping genes (*ace*, *gltB*, *gmhD*, *lepA*, *lipA*, *narK*, *ndh*). The primer sets used for PCR amplification were obtained from *B. pseudomallei* MLST scheme as described in PubMLST ([https://pubmlst.org/bpseudomallei/](https://pubmlst.org/bpseudomallei/)). Sequencing PCR using the same primer set was performed using the Big Dye Terminator (v3.1) cycle sequencing kit (Applied Biosystems, Thermo Fisher Scientific Company, Waltham, MA.) under the manufacturer’s protocol, purified and resolved on ABI 3500 Genetic Analyzer (Applied Biosystems, Thermo Fisher Scientific Company, Waltham, MA). The complete sequences of the seven loci of the house keeping genes were assigned allelic number and defined a sequence type based on the allelic profile match on the PubMLST database. New allele numbers and STs were assigned to sequences not reported previously by submission to the database. Genetic relatedness of the isolates in comparison to the global isolates was analysed using the goeBURST algorithm of the PHYLOVIZ open source software to establish a clonal association. The nucleotide diversity was calculated using the allelic sequences on the DNA sequence polymorphism software (v6.10.01). SplitsTree4 (version 4.14.6) was used to derive a comparative phylogenetic relationship among the study isolates and isolates previously reported from India ([http://www.ub.edu/dnasp/](http://www.ub.edu/dnasp/)).

Ethics statement

The *B. pseudomallei* are from clinical specimens (blood, pus, tissue and urine) from patients admitted to the Christian Medical College and Hospital, Vellore. This study was approved with ethical clearance to use the clinical isolates by the Institutional Review Board with IRB MIN 16044. The clinical samples are anonymized.

Results

Within the thirty one clinical isolates obtained during 2015 to 2016, Systemic forms of melioidosis (blood) contributed to 51.6% (*n* = 16) and localized to 45% (*n* = 14) with one urinary tract infection (3.2%). In the case of the localised infections, the number of isolates from pus and the urinary tract infection were 29% (*n* = 9), tissue 16% (*n* = 5) respectively. Infected patients were from the states of Tamilnadu (*n* = 11; 35.5%), West Bengal (*n* = 5; 16.1%), Andhra Pradesh (*n* = 6; 19.5%), Jharkhand (*n* = 5; 16.1%), Kerala (*n* = 1; 3.2%) and Tripura (*n* = 1; 32%). Twenty isolates had distinct allelic profiles from the existing database and were assigned new sequence types (ST) (13 new STs, ST1630-ST1642). Eleven isolates in this study were found to be previously reported STs (ST51, ST1364, ST1099, ST300, ST1552, ST375, ST56, ST71, ST228 and ST99).

The genetic diversity among the identified STs was low as four isolates had ST1630 (12.9%), three had ST1639 (9.7%), two had ST51, ST1637, 1641 (6.5%) and the remaining 19 isolates having different STs. Isolate identifiers with geographical location, type of specimen and the corresponding STs are represented in Table 1. Interestingly in two patients with both systemic and localized infections (VBBP002 –Systemic–ST1630 and VBBP006 –localized—ST1631; VBBP011—Systemic–ST51 and VBBP023—localized–ST375), different STs were observed with respect to the site of isolation—they are single loci variants. goeBURSTanalysis clustered 18 of the study STs into a single major clonal complex of founder ST300. Four STs (1634, 1632, 1636, and 1639) were found to be singletons and are outliers (Fig 1).
Thirty five percent (n = 11) of the identified STs in this study have been previously reported and were found to be associated with Singapore (ST51), China (ST51, ST1099), Thailand (ST51, ST99, ST375, ST228, ST300), Malaysia (ST51, ST99), Burma (ST51), Bangladesh (ST56), Cambodia (ST56), Vietnam (ST56), Philippines (ST99) and Sri Lanka (ST1364) of Southeast Asia (Fig 2) [9, 10].

Except for the isolates with the ST51 (6.5%) and ST 56 (3%), no association was found between the epidemiological year and the prevalence of the isolates. ST51 and ST 56 were first observed in 1935 and 1947 and are still seen in clinical cases (Table 2).

Nucleotide diversity among the study isolates as calculated by DNA SP6 was 0.00212 and within the Indian isolates was 0.00182 (Table 3). Splits tree analysis depicts 80% of the isolates associated with Southeast Asia into one group wherein the rest of the study isolates are grouped differently (Fig 3).

### Table 1. Details of the study isolates identifiers and corresponding year of isolation, type of infection, geographical location, and the sequence type information.

<table>
<thead>
<tr>
<th>Isolate</th>
<th>Year</th>
<th>Type of infection</th>
<th>Type of Specimen</th>
<th>Patient Location</th>
<th>Sequence type</th>
</tr>
</thead>
<tbody>
<tr>
<td>VBBP001</td>
<td>2016</td>
<td>Localized</td>
<td>Pus</td>
<td>West Bengal</td>
<td>1639</td>
</tr>
<tr>
<td>VBBP002</td>
<td>2016</td>
<td>Systemic</td>
<td>Blood</td>
<td>Tamilnadu</td>
<td>1630</td>
</tr>
<tr>
<td>VBBP003</td>
<td>2016</td>
<td>Systemic</td>
<td>Blood</td>
<td>Tamilnadu</td>
<td>1635</td>
</tr>
<tr>
<td>VBBP004</td>
<td>2016</td>
<td>Systemic</td>
<td>Blood</td>
<td>Jharkhand</td>
<td>1632</td>
</tr>
<tr>
<td>VBBP005</td>
<td>2016</td>
<td>Localized</td>
<td>Urine</td>
<td>West Bengal</td>
<td>1633</td>
</tr>
<tr>
<td>VBBP006</td>
<td>2016</td>
<td>Localized</td>
<td>Pus</td>
<td>Tamilnadu</td>
<td>1631</td>
</tr>
<tr>
<td>VBBP007</td>
<td>2015</td>
<td>Systemic</td>
<td>Blood</td>
<td>Andhra Pradesh</td>
<td>1636</td>
</tr>
<tr>
<td>VBBP008</td>
<td>2015</td>
<td>Systemic</td>
<td>Blood</td>
<td>Andhra Pradesh</td>
<td>1634</td>
</tr>
<tr>
<td>VBBP009</td>
<td>2015</td>
<td>Systemic</td>
<td>Blood</td>
<td>Jharkhand</td>
<td>1637</td>
</tr>
<tr>
<td>VBBP010</td>
<td>2015</td>
<td>Systemic</td>
<td>Blood</td>
<td>Jharkhand</td>
<td>1638</td>
</tr>
<tr>
<td>VBBP011</td>
<td>2016</td>
<td>Systemic</td>
<td>Blood</td>
<td>Tamilnadu</td>
<td>51</td>
</tr>
<tr>
<td>VBBP012</td>
<td>2016</td>
<td>Systemic</td>
<td>Blood</td>
<td>Tamilnadu</td>
<td>51</td>
</tr>
<tr>
<td>VBBP014</td>
<td>2016</td>
<td>Localized</td>
<td>Pus</td>
<td>Andhra Pradesh</td>
<td>1630</td>
</tr>
<tr>
<td>VBBP015</td>
<td>2015</td>
<td>Localized</td>
<td>Tissue</td>
<td>Kerala</td>
<td>1364</td>
</tr>
<tr>
<td>VBBP016</td>
<td>2016</td>
<td>Systemic</td>
<td>Blood</td>
<td>Jharkhand</td>
<td>1099</td>
</tr>
<tr>
<td>VBBP017</td>
<td>2016</td>
<td>Systemic</td>
<td>Blood</td>
<td>Tamilnadu</td>
<td>1630</td>
</tr>
<tr>
<td>VBBP018</td>
<td>2016</td>
<td>Systemic</td>
<td>Blood</td>
<td>Tamilnadu</td>
<td>1639</td>
</tr>
<tr>
<td>VBBP019</td>
<td>2016</td>
<td>Localized</td>
<td>Pus</td>
<td>West Bengal</td>
<td>1639</td>
</tr>
<tr>
<td>VBBP020</td>
<td>2016</td>
<td>Localized</td>
<td>Pus</td>
<td>West Bengal</td>
<td>500</td>
</tr>
<tr>
<td>VBBP021</td>
<td>2016</td>
<td>Localized</td>
<td>Tissue</td>
<td>Tamilnadu</td>
<td>1630</td>
</tr>
<tr>
<td>VBBP022</td>
<td>2016</td>
<td>Localized</td>
<td>Tissue</td>
<td>Tamilnadu</td>
<td>1552</td>
</tr>
<tr>
<td>VBBP023</td>
<td>2016</td>
<td>Localized</td>
<td>Tissue</td>
<td>Tamilnadu</td>
<td>375</td>
</tr>
<tr>
<td>VBBP024</td>
<td>2015</td>
<td>Localized</td>
<td>Pus</td>
<td>Jharkhand</td>
<td>1637</td>
</tr>
<tr>
<td>VBBP025</td>
<td>2015</td>
<td>Localized</td>
<td>Pus</td>
<td>West Bengal</td>
<td>56</td>
</tr>
<tr>
<td>VBBP026</td>
<td>2015</td>
<td>Localized</td>
<td>Pus</td>
<td>Tripura</td>
<td>71</td>
</tr>
<tr>
<td>VBBP027</td>
<td>2015</td>
<td>Systemic</td>
<td>Blood</td>
<td>West Bengal</td>
<td>228</td>
</tr>
<tr>
<td>VBBP028</td>
<td>2015</td>
<td>Localized</td>
<td>Tissue</td>
<td>Bangladesh</td>
<td>99</td>
</tr>
<tr>
<td>VBBP029</td>
<td>2016</td>
<td>Systemic</td>
<td>Blood</td>
<td>Andhra Pradesh</td>
<td>1640</td>
</tr>
<tr>
<td>VBBP030</td>
<td>2016</td>
<td>Systemic</td>
<td>Blood</td>
<td>Andhra Pradesh</td>
<td>1641</td>
</tr>
<tr>
<td>VBBP031</td>
<td>2016</td>
<td>Systemic</td>
<td>Blood</td>
<td>Tamilnadu</td>
<td>1642</td>
</tr>
<tr>
<td>VBBP032</td>
<td>2016</td>
<td>Localized</td>
<td>Pus</td>
<td>Andhra Pradesh</td>
<td>1641</td>
</tr>
</tbody>
</table>

* are the novel sequence types identified in this study.

https://doi.org/10.1371/journal.pntd.0006558.t001
Discussion

The burden of *B. pseudomallei* in India and across the world is of a great concern due to its wide distribution in the community as an environmental etiological agent. Molecular typing by MLST method serves as a powerful epidemiological tool to determine the source of infection (local epidemiology) and understand the diversity and evolution of the pathogen population. Though this study involved a small number of isolates (n = 31), the identified STs in this study provide information on regional and country wide sequence diversity.

The different types of melioidosis in the Southeast Asian region include bacteraemia, skin/soft tissue infections, localized abscesses (spleenic, prostatic, liver, prostatic, parotid), pneumonia and genitourinary tract infections [11]. There was no association found between the different types of melioidosis and the sequence types among the study isolates, which showed high diversity.

Thirty five percent (n = 11) of the study isolates are confined to STs of South East Asia conferring a regional association and the remaining are novel STs. Though there is high diversity among *B. pseudomallei* across India and South East Asia, this study provides insights into the regional STs corresponding to both systemic and localized infections being consistent over a long period of time. The correlation between a few of the identified STs (51, 56) and the epidemiological years denote persistent strains causing infection across the continent.

The variation among isolates (VBBP002:VBBP006 and VBBP011:VBBP023) in both systemic and localized infections from the same patient with a single loci variation shows possible evolution over a short period of time; however genome wide studies are needed to provide valid information. Nucleotide diversity and splits network analysis within the house keeping genes shows the least differences for the 6 housekeeping genes, reducing the possibilities of
recombination events but more of single nucleotide polymorphisms. The nucleotide diversity of the gmhD gene was found to be (0.00306) and has the maximum number of allelic profiles within the study population. The existence of the parallelogram type of split tree network of all the south Indian isolates signifies the possibility of recombination events, but the Southeast Asian study isolates did not show a typical parallelogram and lie on the same group conferring the absence of recombination events [12].

Though Multi locus sequence typing (MLST) is one of the most commonly used Single-Nucleotide-Polymorphism (SNP) based phylogeny with the use of seven housekeeping genes, it represents only 0.05% (3401 nucleotides) of the total bacterial genome of 7.4 million

![Fig 2. Distribution and association of the study ST's with the ST’s from Southeast Asian region retrieved from the PubMLST database (10). Light Green—ST1364 (Kerala & Sri Lanka); Dark Green ST375 (Tamilnadu & Thailand); Black—ST1552 (Tamilnadu and Pondicherry); Red—ST51 (Tamilnadu, Singapore, China, Thailand, Malaysia and Burma); Purple—ST228 (West Bengal, Thailand &Vietnam); Blue—ST1099 (Jharkhand & China); Brown—ST56 (West Bengal Bangladesh, Cambodia & Vietnam); Orange—ST300 (West Bengal and Thailand); Yellow—ST99 (Bangladesh, Philippines, Thailand and Malaysia). The figure was recreated using open source:https://commons.wikimedia.org/wiki/Atlas_of_the_world.](https://doi.org/10.1371/journal.pntd.0006558.g002)

Table 2. Epidemiological year of isolation of the STs identified in the study isolates from different South Asian countries.

<table>
<thead>
<tr>
<th>Sequence Type</th>
<th>Country</th>
<th>Continent</th>
<th>Year of isolation</th>
</tr>
</thead>
<tbody>
<tr>
<td>ST51</td>
<td>Singapore, China, Thailand, Malaysia, Burma</td>
<td>Asia</td>
<td>1935–2015</td>
</tr>
<tr>
<td>ST71</td>
<td>Unknown</td>
<td>Asia</td>
<td>1999, 2000</td>
</tr>
<tr>
<td>ST375</td>
<td>Thailand</td>
<td>Asia</td>
<td>1966</td>
</tr>
<tr>
<td>ST1552</td>
<td>Pondicherry, India</td>
<td>Asia</td>
<td>Unknown</td>
</tr>
<tr>
<td>ST1364</td>
<td>Sri Lanka</td>
<td>Asia</td>
<td>2015</td>
</tr>
<tr>
<td>ST1099</td>
<td>China</td>
<td>Asia</td>
<td>2011</td>
</tr>
</tbody>
</table>

https://doi.org/10.1371/journal.pntd.0006558.t002
### Table 3. Comparison between the nucleotide diversity of MLST housekeeping genes of the study isolates.

<table>
<thead>
<tr>
<th>Gene</th>
<th>Size in bp</th>
<th>Alleles</th>
<th>No of Polymorphic sites</th>
<th>Nucleotide Diversity</th>
</tr>
</thead>
<tbody>
<tr>
<td>ace</td>
<td>519</td>
<td>3</td>
<td>2</td>
<td>0.00142</td>
</tr>
<tr>
<td>gltB</td>
<td>522</td>
<td>5</td>
<td>5</td>
<td>0.00224</td>
</tr>
<tr>
<td>gmhD</td>
<td>468</td>
<td>7</td>
<td>5</td>
<td>0.00306</td>
</tr>
<tr>
<td>lepA</td>
<td>486</td>
<td>4</td>
<td>3</td>
<td>0.00190</td>
</tr>
<tr>
<td>lipA</td>
<td>402</td>
<td>3</td>
<td>3</td>
<td>0.00155</td>
</tr>
<tr>
<td>narK</td>
<td>561</td>
<td>6</td>
<td>7</td>
<td>0.00318</td>
</tr>
<tr>
<td>ndh</td>
<td>443</td>
<td>2</td>
<td>1</td>
<td>0.00118</td>
</tr>
<tr>
<td>Concatenated Sequences</td>
<td>3401</td>
<td>23 STs</td>
<td>26</td>
<td>0.00212</td>
</tr>
<tr>
<td>Concatenated Indian Isolates</td>
<td>3401</td>
<td>54 STs</td>
<td>28</td>
<td>0.00182</td>
</tr>
</tbody>
</table>

https://doi.org/10.1371/journal.pntd.0006558.t003

**Fig 3.** Splits tree of the concatenated allelic sequences of study isolates and previously reported South Indian isolates. STs within the square represents the STs associated to South East Asia reported in the study. It is an unrooted tree with bootstrap value of 1000.

https://doi.org/10.1371/journal.pntd.0006558.g003
nucleotides, with less discriminatory potential being the main limitation in a closely related group or within the sequence type. However, whole genome based SNP typing provides phylogeny with high discriminatory power, which could further type the isolates belonging to same sequence types and/or clonal group. This was substantiated by the studies done by Price et al., 2015, to show the differences between same sequence types but polyclonal by Whole Genome Sequencing in a patient with chronic melioidosis across years unveiling the genome plasticity [13], but the study did not indicate the non synonymous nucleotide polymorphisms. Additionally, Chapple et al 2016 describe the differences in B. pseudomallei by Whole Genome Sequences within the same sequence types being persistent across years and different regions, but the mutations not correlating to the environment factors [14]. This evidence gives a glimpse of the high evolution in B. pseudomallei, with conserving the core genome having strong ancestral relationships as derived in this study.

Prospective studies based on whole genome phylogeny would provide higher resolution over the genome plasticity of B. pseudomallei in India and the regional association through the conserved regions on this pathogen. Continent wide large scale genomic studies would enable us to establish a regional association of the strains [15]. To conclude, future studies must focus on whole genome based SNP typing in order to understand the phylogeny and evolution of this bacterium.

Author Contributions
Conceptualization: Veeraraghavan Balaji, Francis Yesurajan Inbanathan, Suresh Kumar Rajamani Sekar.
Data curation: Veeraraghavan Balaji, Susmitha Perumalla, Rajamani Perumal, Francis Yesurajan Inbanathan, Suresh Kumar Rajamani Sekar, Rani Diana Sahni, John Antony Jude Prakash, Ramya Iyadurai.
Formal analysis: Francis Yesurajan Inbanathan, Suresh Kumar Rajamani Sekar.
Funding acquisition: Veeraraghavan Balaji.
Investigation: Veeraraghavan Balaji, Rani Diana Sahni.
Methodology: Veeraraghavan Balaji, Francis Yesurajan Inbanathan, Suresh Kumar Rajamani Sekar, Miracle Magdelene Paul, Ramya Iyadurai.
Project administration: Veeraraghavan Balaji.
Resources: Veeraraghavan Balaji.
Software: Francis Yesurajan Inbanathan.
Supervision: Susmitha Perumalla.
Validation: Veeraraghavan Balaji, Francis Yesurajan Inbanathan.
Visualization: Veeraraghavan Balaji, Francis Yesurajan Inbanathan.
Writing – original draft: Veeraraghavan Balaji, Francis Yesurajan Inbanathan, Suresh Kumar Rajamani Sekar.
Writing – review & editing: Veeraraghavan Balaji, Susmitha Perumalla, Rajamani Perumal, Suresh Kumar Rajamani Sekar, John Antony Jude Prakash.

References


10. This publication made use of the *Burkholderia pseudomallei* MLST website (https://pubmlst.org/bpseudomallei/) sited at the University of Oxford (Jolley & Maiden 2010, *BMC Bioinformatics*, 11:595). The development of this site has been funded by the Wellcome Trust, October 2017.


