Title: High diversity of Blastocystis subtypes isolated from asymptomatic adults living in Chiang Rai, Thailand

Article Type: Research paper

Keywords: Blastocystis; genetic diversity; prevalence; subtyping; Thailand

Abstract: Blastocystis is a common and broadly distributed microbial eukaryote inhabiting the gut of humans and other animals. The genetic diversity of Blastocystis is extremely high comprising no less than 17 subtypes in mammals and birds. Nonetheless, little is known about the prevalence and distribution of Blastocystis subtypes colonising humans in Thailand. Molecular surveys of Blastocystis remain extremely limited and usually focus on the central, urban part of the country. To address this knowledge gap, we collected stool samples from a population of Thai adults (n=178) residing in Chiang Rai Province. The barcoding region of the small subunit ribosomal RNA was employed to screen for Blastocystis and identify the subtype. Forty-one stool samples (23%) were identified as Blastocystis positive. Six of the nine subtypes that colonise humans were detected with subtype (ST) three being the most common (68%), followed by ST1 (17%) and ST7 (7%). Comparison of subtype prevalence across Thailand using all publicly available sequences showed that subtype distribution differs among geographic regions in the country. ST1 was most commonly encountered in the central region of Thailand, while ST3 dominated in the more rural north and northeast regions. ST2 was absent in the northeast, while ST7 was not found in the center. Thus, this study shows that ST prevalence and distribution differs not only among countries, but also among geographic regions within a country. Potential explanations for these observations are discussed herewith.
April 8, 2018

Editorial Board,
Infection Genetics and Evolution

Dear Editors,

Here we submit a revised version of our manuscript entitled: “High diversity of Blastocystis subtypes isolated from asymptomatic adults living in Chiang Rai, Thailand”.

We have addressed all the comments from the reviewers and hope the manuscript is now suitable for publication.

The manuscript contains only original data that have not been submitted or published elsewhere. We look forward to hearing from you regarding our revised manuscript.

Sincerely,

Eleni Gentekaki, Ph.D.
We thank the reviewers for the positive comments on our manuscript. Below, we address the reviewers’ points. Replies are in bold lettering.

**Reviewer #1**

I think the title is not fit to this manuscript, because the new data obtained in this study was limited from northern region of Thailand.

**We have modified the title as follows:** High diversity of *Blastocystis* subtypes isolated from asymptomatic adults living in Chiang Rai, Thailand

Line 51-:
Description in this sentence is incorrect, because there are many *Blastocystis* organisms had been identified in addition to 17 kinds of the STs.

**We have added the following sentence to account for the lineages composed of reptilian, amphibian and insect sequences and which have not been assigned to subtypes.** Lines 54-56: *Isolates from avian and mammalian hosts* are currently classified into 17 subtypes (STs) based on the divergence of the small subunit ribosomal RNA (SSU rRNA), though new STs might be arising (Alfellani *et al.*, 2013c; Betts *et al.*, 2018;). *Blastocystis* isolated from reptilian, amphibian and insect hosts are distinct and are not assigned into subtypes.

Lines 56-:
I think that ST9 is only isolated from humans.

**We have added the following sentence in lines 59-60:** The exception is ST9, which has yet to be found in a non-human host.

Line 153-:
I am confusing "the alignment contained 1,379 sites", because only barcode-region (ca. 600bp) was amplified by the nested-PCR for sequencing in this study.

**Indeed, even though we amplified the barcode region, our alignment contained many full-length sequences (~1800bp).** Thus, after trimming, the remaining sites were more than 600bp. This is standard practice in phylogenetics, whereby missing data is the norm. By doing this we improve resolution of the phylogenetic tree and strengthen support at the nodes. The topology of the tree remains unchanged. Many investigators do the same. For instance, Alfellani *et al.*, 2013. Genetic diversity of Blastocystis in Livestock and zoo animals, also amplified the barcoding region, but their alignment contained 1,462 sites.
Discussion
I feel the discussion is too long for the results, please concise the discussion. Several parts may be deleted where doesn't match the purpose of this study (lines 83-).

Our discussion consists of four parts: 1) asymptomatic carriage of Blastocystis, 2) comparison of prevalence rates to other studies in Thailand, 3) subtypes detected in this study, and 4) subtype distribution in Thailand. We have shortened the discussion in parts three and four.

Reviewer #2

My only comment is that it is a small study sample number for the Chiang Rai study population and all are asymptomatic. This makes it difficult to know the true prevalence of Blastocystis in this region. Only using asymptomatic people means that you might be missing a large number of positives and potential different STs which may come from symptomatic people. This limitation should be stated in the discussion.

We agree with the reviewer’s comment. The purpose of this pilot study was to examine which subtypes circulate in asymptomatic individuals. To that end, we have changed the title to reflect our study site and sampled population and we have also emphasized this in the discussion (Lines: 204-205, 270-271, 278-279).
KEY FINDINGS
- First study of Blastocystis prevalence and subtyping in Chiang Rai Province, Thailand
- Six of the nine subtypes known to colonize humans were found
- ST3 is the dominant ST followed by ST1 and ST7
- Distribution of Blastocystis subtypes among geographic regions of Thailand differs
High diversity of *Blastocystis* subtypes isolated from asymptomatic adults living in Chiang Rai, Thailand

Amara Yowang\(^a\), Anastasios D. Tsaousis\(^b\), Tawatchai Chumphonsuk\(^a\), Nontaphat Thongsin\(^a\), Niwed Kullawong\(^c,d\), Siam Popluechai\(^a,d\), Eleni Gentekaki\(^\ast\)\(^h\)\(^*\)

\(^a\) School of Science, Mae Fah Luang University, Chiang Rai, Thailand

\(^b\) Laboratory of Molecular and Evolutionary Parasitology, RAPID group, School of Biosciences, University of Kent, Canterbury, Kent, UK

\(^c\) School of Health Science, Mae Fah Luang University, Chiang Rai, Thailand

\(^d\) Human Gut Microbiome for Health Research Unit, Mae Fah Luang University, Chiang Rai, Thailand

Running title: *Blastocystis* diversity and subtyping in Thai adults

\(^\ast\) Corresponding author: Eleni Gentekaki, Mae Fah Luang University, Chiang Rai, Thailand. Tel: +66 5391 6776, Fax: +66 5391 6768, E-mail: gentekaki.ele@mfu.ac.th
Abstract

*Blastocystis* is a common and broadly distributed microbial eukaryote inhabiting the gut of humans and other animals. The genetic diversity of *Blastocystis* is extremely high comprising no less than 17 subtypes in mammals and birds. Nonetheless, little is known about the prevalence and distribution of *Blastocystis* subtypes colonising humans in Thailand. Molecular surveys of *Blastocystis* remain extremely limited and usually focus on the central, urban part of the country. To address this knowledge gap, we collected stool samples from a population of Thai adults (n=178) residing in Chiang Rai Province. The barcoding region of the small subunit ribosomal RNA was employed to screen for *Blastocystis* and identify the subtype. Forty-one stool samples (23%) were identified as *Blastocystis* positive. Six of the nine subtypes that colonise humans were detected with subtype (ST) three being the most common (68%), followed by ST1 (17%) and ST7 (7%). Comparison of subtype prevalence across Thailand using all publicly available sequences showed that subtype distribution differs among geographic regions in the country. ST1 was most commonly encountered in the central region of Thailand, while ST3 dominated in the more rural north and northeast regions. ST2 was absent in the northeast, while ST7 was not found in the center. Thus, this study shows that ST prevalence and distribution differs not only among countries, but also among geographic regions within a country. Potential explanations for these observations are discussed herewith.

Keywords: *Blastocystis*; genetic diversity; prevalence; subtyping; Thailand
1. Introduction

*Blastocystis* is the most commonly found protist in the gastrointestinal tract of humans and other animals (Roberts *et al.*, 2013; Tan, 2008). Though *Blastocystis* exhibits morphological stasis, it has a high degree of genetic heterogeneity. Isolates from avian and mammalian hosts are currently classified into 17 subtypes (STs) based on the divergence of the small subunit ribosomal RNA (SSU rRNA), though new STs might be arising (Alfellani *et al.*, 2013c; Betts *et al.*, 2018). *Blastocystis* isolated from reptilian, amphibian and insect hosts are distinct and are not assigned into subtypes.

So far, only ST1-ST9 have been found in humans with ST1-ST4 being the most common colonisers (Alfellani *et al.*, 2013b). However, these STs have also been detected in a variety of non-human hosts, suggesting that *Blastocystis* has low host specificity (Stensvold and Clark 2016). The exception is ST9, which has yet to be found in a non-human host. The rest of the STs have so far been identified in both domesticated and wildlife animal hosts (Alfellani *et al.*, 2013a).

Lately, considerable research effort has gone into deciphering potential roles of *Blastocystis* in intestinal disease. For instance, an association between presence of *Blastocystis* and irritable bowel syndrome has been discussed frequently, though a definitive causal link has yet to be postulated (Jimenez-Gonzalez *et al.*, 2012; Khademvatan *et al.*, 2017; Poirer *et al.*, 2012; Yakoob *et al.*, 2010). Nonetheless, emerging data have shown presence of *Blastocystis* in asymptomatic individuals leading to the hypothesis that the organism is part of the healthy adult gut microbiota and that only specific STs and/or genotypes might be pathogenic (Chabe *et al.*, 2017; Scanlan and Stensvold, 2013; Scanlan *et al.*, 2014; Stensvold and van der Giezen, 2018). In this context, surveys examining prevalence of *Blastocystis* should ideally be
accompanied by subtyping. Thus, while the cosmopolitan status of Blastocystis is well established, geographic distribution of specific subtypes and their prevalence are relatively recent endeavors.

Research efforts on the latter two are disproportionate among countries and/or geographic regions: a plethora of Blastocystis studies come from the Americas and Europe, whilst Asia remains disproportionately sampled, given that half of the earth’s population resides in Asian countries (El Safadi et al., 2016; Scanlan et al., 2016; Villegas-Gomez et al., 2016). Collectively, these studies have shown that ST3 is the most prevalent one in almost all regions that have been examined, with a notable exception (Oliveira-Arbex et al., 2018). Prevalence of the other common STs – namely ST1, ST2, and ST4 – seems to be dependent on geographic region, while ST6-ST9 are not as commonly encountered (Beghini et al., 2017; Stensvold and Clark, 2016).

Herein, we collected samples from 178 asymptomatic adult humans living in Chiang Rai Province, Thailand and examined the prevalence, genetic diversity and distribution of Blastocystis. The region combines several particularities, which have created a culture and lifestyle that differs from other regions of Thailand. We provide the first subtyping data from the area, thus expanding knowledge on Blastocystis ST distribution and prevalence in the country.

2. Materials and Methods

2.1. Ethics statement

The human ethics committee of Mae Fah Luang University approved collection of fecal samples from adult Thai volunteers (License approval number
Volunteers signed an informed consent form before participating in the study.

2.2. Human subjects

All volunteers were over 18 years old, Thai and lived in Chiang Rai Province at the time of sampling (Fig. 1). In total, 178 adult volunteers participated in the study. The volunteers had no history of gastrointestinal symptoms, no diarrhea episodes a month prior to sampling and had received no antibiotic treatment at least two months prior to sample collection.

2.3. Stool collection and DNA extraction

Fecal samples were collected from the volunteers using sealed, sterile containers and stored at -80°C until DNA extraction. Samples were collected from six districts of Chiang Rai Province: Mae Chan (n=20), Mae Lao (n=20), Mae Sai (n=25), Muang (n=69), Pa Daet (n=23) and Phaya Mengrai (n=21) (Fig. 1). Total DNA was isolated using the QIAamp DNA Stool Mini Kit (Qiagen, Thailand) according to manufacturer’s specifications. The extracted DNA was stored at -20 °C for later use.

2.4. PCR amplification, amplicon purification and sequencing

Polymerase chain reaction (PCR) was used to amplify a 650 bp fragment of the SSUrRNA at the 5’ of the gene. This fragment is considered the barcoding region of Blastocystis spp. (Scicluna et al., 2006).

Two PCR reactions were used to amplify the SSUrRNA gene of Blastocystis. The broadly specific primers, RD5 5’ –
GGAAGCTTATCTGGTTGATCCTGCCAGTA – 3’ and RD3 5’ –
GGGATCCTGATCCTTCCGCAGGTTCACCTAC – 3’ were used for the primary PCR reaction (Clark, 1997). The total volume of each reaction was 25μl containing the following reagents: 10X reaction buffer, 1 mM MgCl₂, 0.1 mM dNTPs, 0.2 μM forward primer, 0.2 μM reverse primer, 5U Taq DNA polymerase (RBC Bioscience, New Taipei City, Taiwan) and 12 μl HPLC grade water. Cycling conditions were as follows: initial denaturation at 94 °C for 3 min followed by 35 cycles of denaturation at 94 °C for 1 min 40 s, annealing at 65 °C for 1 min 40 s, extension at 72 °C for 1 min 40 s followed by a final extension at 72 °C for 10 min.

The more narrowly specific primers RD5F 5’ – ATCTGGTCGATCCTGCCAGT – 3’ and BhRDr 5’ – GAGCTTTTTAACCTGCAACAACG – 3’ were employed for the secondary, nested PCR reaction. Reaction volume and concentration of reagents were as in the primary PCR with the exception of the DNA template. Instead of genomic fecal DNA, one μl from the primary PCR reaction was used. Cycling conditions were as follows: denaturation at 94 °C for 3 min, followed by 35 cycles of denaturation at 94 °C for 1 min, annealing at 55 °C for 1 min and extension at 72 °C for 1 min and a final extension at 72 °C for 10 min.

Positive PCR reactions were purified using the HiYield™ Gel/PCR Fragments Extraction Kit (RBC bioscience, USA) according to manufacturer’s specifications. Purified PCR products were sequenced by 1stBASE, Malaysia.

2.5. Subtyping and phylogenetic analysis

The obtained sequences were examined and poor quality bases at the 5’ and 3’ ends of the sequence were removed. Sequences were then subjected to BLAST search against GenBank database to exclude contamination. Sequences identified as
Blastocystis, were further checked against the Blastocystis Subtype Sequence Typing (MLST) publicly available at http://pubmlst.org/blastocystis/, in order to identify subtype and the corresponding allele.

The new sequences identified as Blastocystis along with sequences spanning the breadth of diversity of Blastocystis (all subtypes and sequences from ectotherms) were used to construct the dataset. In total, 103 sequences were aligned using MAFFT v. 7 (Katoh and Toh, 2010), visually inspected and further improved manually. Ambiguous regions were removed with trimAl v. 1.3 (Capella-Gutierrez et al., 2009). After trimming, the alignment contained 1,379 sites. A maximum likelihood (ML) phylogenetic tree was constructed using RAxML v. 8 (Stamatakis, 2006) and the general time reversible + \( \Gamma \) model of nucleotide substitution. The heuristic tree search was based on 20 starting trees. One thousand bootstrap replicates were generated and used for ML analysis. MrBayes v. 3.2.6 (Ronquist and Huelsenbeck, 2003) was used to construct the Bayesian Inference (BI) tree using the same model as above. Two sets of four independent Markov Chain Monte Carlo simulations were run for 1,500,000 generations. Sampling was done every 1,000 generations and 25% were discarded as burn-in. Convergence was declared when the standard deviation of split sequences was less than 0.01. Phylogenetic analyses were conducted on the Cipres Science Gateway (http://www.phylo.org/portal2/home.action).

To further examine distribution of subtypes across Thailand, we downloaded all Blastocystis sequences with the following identifiers: country = Thailand and host = Homo sapiens. In total, 161 Blastocystis sequences were used for the analysis of ST prevalence. For the ST distribution across Thailand 159 were used. Two sequences were removed because the exact locality of collection was not noted.
3. Results

3.1. Screening of fecal samples

Of the 178 samples, 41 (23%) were sequence positive for *Blastocystis* (Table 1). Differences were noted in the percent prevalence among districts. The highest prevalence was observed in Phaya Mengrai District (33%), while the lowest was noted in Mae Chan and Mae Lao Districts (15%).

3.2. Subtype prevalence and distribution

*Blastocystis* prevalence in our study population was 23% (41/178). In total, six of the nine STs that colonise humans were identified in this study: ST1 (n=7, 17%), ST2 (n=1, 2%), ST3 (n=28, 68%), ST4 (n=1, 2%), ST6 (n=1, 2%) and ST7 (n=3, 7%). Two alleles of ST1 were detected with the dominant one being allele 4, while four alleles were found within the dominant ST3 (Fig. 2). Even though only three sequences are available for ST7, two alleles were detected, suggesting a high degree of genetic diversity for this ST in the study population. The prevalence of the various *Blastocystis* subtypes in Thailand resembles the rest of the world (Fig. 3). Subtype 3 is the most prevalent one, followed by ST1, ST2, ST6 and ST7. Nonetheless, notable differences emerge when looking at the prevalence of STs according to geographic region (Fig. 4).

All newly generated sequences have been submitted to GenBank (submission number SUB3865516)

3.3. Phylogenetic analyses

All newly derived sequences grouped within clades formed by previously reported STs (Fig. 5). In agreement with previous studies, isolates from ectothermic
metazoans and those from ST15, ST16 and ST17 placed at a basal position (Betts et al., 2018; Yoshikawa et al., 2016). Placement of the new sequences confirmed ST predictions from the pubmlst website. The rest of the STs grouped into two large clusters, once consisting of ST3, ST4, ST8 and ST10 and the other formed by the remaining STs.

4. Discussion

All samples herein were collected from asymptomatic adults with no history of gastrointestinal disease. Our results match previous recent reports, which have demonstrated asymptomatic carriage of Blastocystis in populations from westernised and non-westernised countries, as well as, rural and urban settings (Lukeš et al., 2015; Parfrey et al., 2014; Scanlan et al., 2014). Findings from analyses of large datasets are also consistent with these observations (Audebert et al., 2016; Beghini et al., 2017). These are in contrast with many previous studies that postulate links of Blastocystis with intestinal diseases and dysbiosis (Poirer et al., 2012; Yakoob et al., 2010). Thus, it is still unclear, whether Blastocystis is a pathogen, a harmless commensal or a beneficial member of gut microbiota. The debate has given rise to the speculation that pathogenicity of Blastocystis is limited to specific genotypes with the majority of them being harmless (Scanlan and Stensvold, 2013; Tan et al., 2010).

Given the polymicrobial nature of the gut, an as yet unexplored theme in the context of ecological theory, is whether Blastocystis becomes pathogenic through synergistic action with other microbes and/or their metabolic products. Nonetheless, the asymptomatic carriage of Blastocystis found in our study and others reinforces the idea of it being a part of the healthy gut microbiota.
We used PCR to amplify the barcoding region of *Blastocystis* (Scicluna et al., 2006) from 178 fecal samples. This method demonstrated that the overall prevalence of *Blastocystis* in our study population was 23%. It is difficult to determine how this prevalence rate compares to other studies from Thailand. *Blastocystis* detection in the country is primarily based on microscopy (Boonjaraspinyo et al., 2013; Kitvatanachai and Rhongbutsri, 2013; Kitvatanachai et al., 2008; Ngrenngarmlert et al., 2007; Pipatsatitpong et al., 2012; Sanguankiat et al., 2014). In general, microscopy is less sensitive and underestimates prevalence. Molecular-based studies on *Blastocystis* constitute a recent development and remain extremely limited (Boondit et al., 2014; Jantermtor et al., 2013; Palasuwan et al., 2016; Pintong et al., 2014; Pipatsatitpong et al., 2015; Popruk et al., 2015; Thathaisong et al., 2013). These molecular surveys have shown prevalence ranging from 6% to 51%. The observed variable data from our study and others could be the result of demographically different populations being surveyed. For instance, the highest abundance of *Blastocystis* was found in orphan male children, while residents of communities along Chao Praya River in Ayutthaya Province (central Thailand) had the lowest prevalence (Palasuwan et al., 2016; Pintong et al., 2014). Age, health status of the individuals, and general lifestyle are likely contributing factors accounting for the observed differences among studies (Leelayoova et al., 2008).

Subtyping of *Blastocystis* positive samples identified six of the nine STs known to colonise humans: ST1, ST2, ST3, ST4, ST6 and ST7. ST3 was dominant and present in two thirds of the samples, followed by ST1 and ST7. This is slightly different than the global trend, whereby ST3, ST1 and ST2, in this order, are the most frequently encountered (Stensvold and Clark, 2016). This pattern is also noted in molecular surveys of *Blastocystis* in neighboring Southeast Asian countries (Laos,
Cambodia and Malaysia), except for Laos, where ST1 was the most frequent followed
by ST3 and then ST2 (Nithyamathi et al., 2016; Noradilah et al., 2017; Sanpool et al.,
2017; Wang et al., 2014). Herein, ST2, ST4 and ST6 were found in a single individual
each. ST4 is a geography-influenced ST and is typically found in Europe, where it is
abundant and highly prevalent (Alfellani et al., 2013b; Beghini et al., 2017; Forsell et
al., 2016). Indeed, ST4 occurrence in Southeast Asia is extremely rare having been
found so far only in Thailand and Malaysia (Nithyamathi et al., 2016; Noradilah et
al., 2017; Popruk et al., 2015).

When all available sequences from human stool samples from Thailand are
taken into account (including ours), the most prevalent STs are ST3, ST1 and ST2, a
pattern that follows the global trend (Stensvold and Clark, 2016). Given the known
influence of geography on ST distribution, we zoomed into specific regions. Based on
geography, Thailand is divided into four regions: north, northeast, center and south.
Thus, when geographic region is taken into account, there are some notable
differences in the ST distribution. For instance, the relative abundance of ST3 and
ST1 changes. ST1 is dominant in the center, while ST3 dominates in the north and
northeast. ST2 is absent in the northeast, whilst ST4 has only been found so far in the
north. Importantly, the center, which is the most urbanised region of Thailand, has the
least number of subtypes. Incidentally, this is also the most broadly surveyed region
and where most sequences come from. Thus, though we cannot exclude the possibility
that some STs circulate in the center and have yet to be detected, this is not very
likely. Presence of ST6 and ST7 in the north and northeast could be attributed to
domestic fowl, especially chickens and geese, which live in close proximity to
humans.
The data obtained herein are the first on prevalence of *Blastocystis* and distribution of STs in asymptomatic individuals living in Chiang Rai Province. Our data along with other molecular data from Thailand will contribute to the knowledge on the epidemiology of *Blastocystis* in the country. Nonetheless, major gaps exist, in that: 1) molecular data on *Blastocystis* from Thailand is scant, thus ST distribution is largely unknown and, 2) the majority of Thai communities have yet to be sampled. For example, there is no molecular data of *Blastocystis* from the south part of Thailand. Moving forward, future studies should expand to generate subtyping information from more communities across Thailand and also include symptomatic individuals.

**CONFLICTS OF INTEREST**

The authors declare no conflict of interest

**ACKNOWLEDGMENTS**

We are grateful to all the volunteers that participated in this study. We also wish to acknowledge the Human Gut Microbiome for Health Research Unit, Mae Fah Luang University, Chiang Rai, Thailand.

**FINANCIAL SUPPORT**

This project was supported by the Thailand Research Fund (RSA6080048) awarded to E.G. A.D.T. was supported by BBSRC research grant (BB/M009971/1).

**REFERENCES**

Alfellani, M.A., Jacob, A.S., Perea, O.N., Krecek, R.C., Taner-Mulla, D., Verweij,


Boonjaraspanyo, S., Boonmars, T., Kaewsamut, B., Ekobol, N., Laummaunwai, P.,


Blastocystis infection is associated with irritable bowel syndrome in a Mexican patient population. Parasitol Res. 110, 1269-1275.


Scanlan, P.D., Stensvold, C.R., Rajilic-Stojanovic, M., Heilig, H.G.H.J., De Vos,
The microbial eukaryote *Blastocystis* is a prevalent and diverse member of the healthy human gut microbiota.


Villegas-Gómez, I., Martínez-Hernández, F., Urrea-Quezada, A., González-Díaz, M.,


FIGURE LEGENDS

Fig. 1. Map of Thailand and Chiang Rai Province districts. Sampling localities used in this investigation are depicted in red lettering.

Fig. 2. Frequency of Blastocystis alleles detected in the study population

Fig. 3. Prevalence of Blastocystis subtypes in Thailand (n=161)

Fig. 4. Prevalence and distribution of Blastocystis subtypes across geographic regions in Thailand

Fig. 5. Maximum likelihood phylogenetic tree inferred from 103 SSUrRNA sequences and 1379 sites. New sequences are depicted in bold lettering. Numerical values on the tree branches indicate bootstrap support percentages and posterior probabilities in this order.
Table 1. Prevalence of *Blastocystis* in six districts in Chiang Rai Province, Thailand

<table>
<thead>
<tr>
<th>District</th>
<th>No of samples</th>
<th>% prevalence (positives)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mae Chan</td>
<td>20</td>
<td>15 (3/20)</td>
</tr>
<tr>
<td>Mae Lao</td>
<td>20</td>
<td>15 (3/20)</td>
</tr>
<tr>
<td>Mae Sai</td>
<td>25</td>
<td>24 (6/25)</td>
</tr>
<tr>
<td>Muang</td>
<td>69</td>
<td>25 (17/69)</td>
</tr>
<tr>
<td>Pa Daet</td>
<td>23</td>
<td>22 (5/23)</td>
</tr>
<tr>
<td>Phaya Mengrai</td>
<td>21</td>
<td>33 (7/21)</td>
</tr>
<tr>
<td>Total</td>
<td>178</td>
<td>23 (41/178)</td>
</tr>
</tbody>
</table>
High diversity and variable geographic distribution of Blastocystis subtypes isolated from asymptomatic adults living in Chiang Rai, Thailand

Amara Yowang\textsuperscript{a}, Anastasios D. Tsaousis\textsuperscript{b}, Tawatchai Chumphonsuk\textsuperscript{a}, Nontaphat Thongsin\textsuperscript{a}, Niwed Kullawong\textsuperscript{c,d}, Siam Popleuechai\textsuperscript{a,d}, Eleni Gentekaki\textsuperscript{i,*}

\textsuperscript{a} School of Science, Mae Fah Luang University, Chiang Rai, Thailand
\textsuperscript{b} Laboratory of Molecular and Evolutionary Parasitology, RAPID group, School of Biosciences, University of Kent, Canterbury, Kent, UK
\textsuperscript{c} School of Health Science, Mae Fah Luang University, Chiang Rai, Thailand
\textsuperscript{d} Human Gut Microbiome for Health Research Unit, Mae Fah Luang University, Chiang Rai, Thailand

Running title: Blastocystis diversity and subtyping in Thai adults

*Corresponding author: Eleni Gentekaki, Mae Fah Luang University, Chiang Rai, Thailand. Tel: +66 5391 6776, Fax: +66 5391 6768, E-mail: gentekaki.ele@mfu.ac.th
Abstract

*Blastocystis* is a common and broadly distributed microbial eukaryote inhabiting the gut of humans and other animals. The genetic diversity of *Blastocystis* is extremely high comprising no less than 17 subtypes in mammals and birds. Nonetheless, little is known about the prevalence and distribution of *Blastocystis* subtypes colonising humans in Thailand. Molecular surveys of *Blastocystis* remain extremely limited and usually focus on the central, urban part of the country. To address this knowledge gap, we collected stool samples from a population of Thai adults (n=178) residing in Chiang Rai Province. The barcoding region of the small subunit ribosomal RNA was employed to screen for *Blastocystis* and identify the subtype. Forty-one stool samples (23%) were identified as *Blastocystis* positive. Six of the nine subtypes that colonise humans were detected with subtype (ST) three being the most common (68%), followed by ST1 (17%) and ST7 (7%). Comparison of subtype prevalence across Thailand using all publicly available sequences showed that subtype distribution differs among geographic regions in the country. ST1 was most commonly encountered in the central region of Thailand, while ST3 dominated in the more rural north and northeast regions. ST2 was absent in the northeast, while ST7 was not found in the center. Thus, this study shows that ST prevalence and distribution differs not only among countries, but also among geographic regions within a country. Potential explanations for these observations are discussed herewith.

Keywords: *Blastocystis*; genetic diversity; prevalence; subtyping; Thailand
1. Introduction

Blastocystis is the most commonly found protist in the gastrointestinal tract of humans and other animals (Roberts et al., 2013; Tan, 2008). Though Blastocystis exhibits morphological stasis, it has a high degree of genetic heterogeneity. Isolates from avian and mammalian hosts are the organism is currently classified into 17 subtypes (STs) based on the divergence of the small subunit ribosomal RNA (SSU rRNA), though new STs might be arising (Alfellani et al., 2013c; Betts et al., 2018;). Blastocystis isolated from reptilian, amphibian and insect hosts are distinct and are not assigned into subtypes. So far, only ST1-ST9 have been found in humans with ST1-ST4 being the most common colonisers (Alfellani et al., 2013b). However, these STs have also been detected in a variety of non-human hosts, suggesting that Blastocystis has low host specificity (Stensvold and Clark 2016). The exception is ST9, which has yet to be found in non-human hosts. The rest of the STs have so far been identified in both domesticated and wildlife animal hosts (Alfellani et al., 2013a).

Lately, considerable research effort has gone into deciphering potential roles of Blastocystis in intestinal disease. For instance, an association between presence of Blastocystis and irritable bowel syndrome has been discussed frequently, though a definitive causal link has yet to be postulated (Jimenez-Gonzalez et al., 2012; Khademvatan et al., 2017; Poirer et al., 2012; Yakoob et al., 2010). Nonetheless, emerging data have shown presence of Blastocystis in asymptomatic individuals leading to the hypothesis that the organism is part of the healthy adult gut microbiota and that only specific STs and/or genotypes might be pathogenic (Chabe et al., 2017; Scanlan and Stensvold, 2013; Scanlan et al., 2014; Stensvold and van der Giezen, 2018). In this context, surveys examining prevalence of Blastocystis should ideally be
accompanied by subtyping. Thus, while the cosmopolitan status of Blastocystis is well established, geographic distribution of specific subtypes and their prevalence are relatively recent endeavors.

Research efforts on the latter two are disproportionate among countries and/or geographic regions: a plethora of Blastocystis studies come from the Americas and Europe, whilst Asia remains disproportionately sampled, given that half of the earth’s population resides in Asian countries (El Safadi et al., 2016; Scanlan et al., 2016; Villegas-Gomez et al., 2016). Collectively, these studies have shown that ST3 is the most prevalent one in almost all regions that have been examined, with a notable exception (Oliveira-Arbex et al., 2018). Prevalence of the other common STs – namely ST1, ST2, and ST4 – seems to be dependent on geographic region, while ST6-ST9 are not as commonly encountered (Beghini et al., 2017; Stensvold and Clark, 2016).

Herein, we collected samples from 178 asymptomatic adult humans living in Chiang Rai Province, Thailand and examined the prevalence, genetic diversity and distribution of Blastocystis. The region combines several particularities, which have created a culture and lifestyle that differs from other regions of Thailand. We provide the first subtyping data from the area, thus expanding knowledge on Blastocystis ST distribution and prevalence in the country.

2. Materials and Methods

2.1. Ethics statement

The human ethics committee of Mae Fah Luang University approved collection of fecal samples from adult Thai volunteers (License approval number
Volunteers signed an informed consent form before participating in the study.  

2.2. Human subjects

All volunteers were over 18 years old, Thai and lived in Chiang Rai Province at the time of sampling (Fig. 1). In total, 178 adult volunteers participated in the study. The volunteers had no history of gastrointestinal symptoms, no diarrhea episodes a month prior to sampling and had received no antibiotic treatment at least two months prior to sample collection.

2.3. Stool collection and DNA extraction

Fecal samples were collected from the volunteers using sealed, sterile containers and stored at -80°C until DNA extraction. Samples were collected from six districts of Chiang Rai Province: Mae Chan (n=20), Mae Lao (n=20), Mae Sai (n=25), Muang (n=69), Pa Daet (n=23) and Phaya Mengrai (n=21) (Fig. 1). Total DNA was isolated using the QIAamp DNA Stool Mini Kit (Qiagen, Thailand) according to manufacturer’s specifications. The extracted DNA was stored at -20°C for later use.

2.4. PCR amplification, amplicon purification and sequencing

Polymerase chain reaction (PCR) was used to amplify a 650 bp fragment of the SSUrRNA at the 5’ of the gene. This fragment is considered the barcoding region of Blastocystis spp. (Scicluna et al., 2006).

Two PCR reactions were used to amplify the SSUrRNA gene of Blastocystis. The broadly specific primers, RD5 5’ –
GGAAGCTTATCTGGTTGATCCTGCCAGTA – 3’ and RD3 5’ –

GGGATCCTGATCCTTCCGCAGGTTCACCTAC – 3’ were used for the primary

PCR reaction (Clark, 1997). The total volume of each reaction was 25μl containing

the following reagents: 10X reaction buffer, 1 mM MgCl₂, 0.1 mM dNTPs, 0.2 μM

forward primer, 0.2 μM reverse primer, 5U Taq DNA polymerase (RBC Bioscience,

New Taipei City, Taiwan) and 12 μl HPLC grade water. Cycling conditions were as

follows: initial denaturation at 94 °C for 3 min followed by 35 cycles of denaturation

at 94 °C for 1 min 40 s, annealing at 65 °C for 1 min 40 s, extension at 72 °C for 1

min 40 s followed by a final extension at 72 °C for 10 min.

The more narrowly specific primers RD5F 5’ – ATCTGGTCGATCCTG

CCAGT – 3’ and BhRDr 5’ – GAGCTTTTTAACTGCAACAACG – 3’ were

employed for the secondary, nested PCR reaction. Reaction volume and concentration

of reagents were as in the primary PCR with the exception of the DNA template.

Instead of genomic fecal DNA, one μl from the primary PCR reaction was used.

Cycling conditions were as follows: denaturation at 94 °C for 3 min, followed by 35

cycles of denaturation at 94 °C for 1 min, annealing at 55 °C for 1 min and extension

at 72 °C for 1 min and a final extension at 72 °C for 10 min.

Positive PCR reactions were purified using the *HiYield™ Gel/PCR Fragments

Extraction Kit* (RBC bioscience, USA) according to manufacturer’s specifications.

Purified PCR products were sequenced by 1stBASE, Malaysia.

2.5. Subtyping and phylogenetic analysis

The obtained sequences were examined and poor quality bases at the 5’ and

3’ ends of the sequence were removed. Sequences were then subjected to BLAST

search against GenBank database to exclude contamination. Sequences identified as
Blastocystis, were further checked against the Blastocystis Subtype Sequence Typing (MLST) publicly available at http://pubmlst.org/blastocystis/, in order to identify subtype and the corresponding allele.

The new sequences identified as Blastocystis along with sequences spanning the breadth of diversity of Blastocystis (all subtypes and sequences from ectotherms) were used to construct the dataset. In total, 103 sequences were aligned using MAFFT v. 7 (Katoh and Toh, 2010), visually inspected and further improved manually. Ambiguous regions were removed with trimAl v. 1.3 (Capella-Gutierrez et al., 2009). After trimming, the alignment contained 1,379 sites. A maximum likelihood (ML) phylogenetic tree was constructed using RAxML v. 8 (Stamatakis, 2006) and the general time reversible + Γ model of nucleotide substitution. The heuristic tree search was based on 20 starting trees. One thousand bootstrap replicates were generated and used for ML analysis. MrBayes v. 3.2.6 (Ronquist and Huelsenbeck, 2003) was used to construct the Bayesian Inference (BI) tree using the same model as above. Two sets of four independent Markov Chain Monte Carlo simulations were run for 1,500,000 generations. Sampling was done every 1,000 generations and 25% were discarded as burn-in. Convergence was declared when the standard deviation of split sequences was less than 0.01. Phylogenetic analyses were conducted on the Cipres Science Gateway (http://www.phylo.org/portal2/home.action).

To further examine distribution of subtypes across Thailand, we downloaded all Blastocystis sequences with the following identifiers: country = Thailand and host = Homo sapiens. In total, 161 Blastocystis sequences were used for the analysis of ST prevalence. For the ST distribution across Thailand 159 were used. Two sequences were removed because the exact locality of collection was not noted.
3. Results

3.1. Screening of fecal samples

Of the 178 samples, 41 (23%) were sequence positive for *Blastocystis* (Table 1). Differences were noted in the percent prevalence among districts. The highest prevalence was observed in Phaya Mengrai District (33%), while the lowest was noted in Mae Chan and Mae Lao Districts (15%).

3.2. Subtype prevalence and distribution

*Blastocystis* prevalence in our study population was 23% (41/178). In total, six of the nine STs that colonize humans were identified in this study: ST1 (n=7, 17%), ST2 (n=1, 2%), ST3 (n=28, 68%), ST4 (n=1, 2%), ST6 (n=1, 2%) and ST7 (n=3, 7%). Two alleles of ST1 were detected with the dominant one being allele 4, while four alleles were found within the dominant ST3 (Fig. 2). Even though only three sequences are available for ST7, two alleles were detected, suggesting a high degree of genetic diversity for this ST in the study population. The prevalence of the various *Blastocystis* subtypes in Thailand resembles the rest of the world (Fig. 3). Subtype 3 is the most prevalent one, followed by ST1, ST2, ST6 and ST7. Nonetheless, notable differences emerge when looking at the prevalence of STs according to geographic region (Fig. 4).

All newly generated sequences have been submitted to GenBank (submission number SUB3865516)

3.3. Phylogenetic analyses

All newly derived sequences grouped within clades formed by previously reported STs (Fig. 5). In agreement with previous studies, isolates from ectothermic
metazoans and those from ST15, ST16 and ST17 placed at a basal position (Betts et al., 2018; Yoshikawa et al., 2016). Placement of the new sequences confirmed ST predictions from the pubmlst website. The rest of the STs grouped into two large clusters, once consisting of ST3, ST4, ST8 and ST10 and the other formed by the remaining STs.

4. Discussion

All samples herein were collected from asymptomatic adults with no history of gastrointestinal disease. Our results match previous recent reports, which have demonstrated asymptomatic carriage of Blastocystis in populations from westernised and non-westernised countries, as well as, rural and urban settings (Lukeš et al., 2015; Parfrey et al., 2014; Scanlan et al., 2014). Findings from analyses of large datasets are also consistent with these observations (Audebert et al., 2016; Beghini et al., 2017). These are in contrast with many previous studies that postulate links of Blastocystis with intestinal diseases and dysbiosis (Poirier et al., 2012; Yakoob et al., 2010). Thus, it is still unclear, whether Blastocystis is a pathogen, a harmless commensal or a beneficial member of gut microbiota. The debate has given rise to the speculation that pathogenicity of Blastocystis is limited to specific genotypes with the majority of them being harmless (Scanlan and Stensvold, 2013; Tan et al., 2010). Given the polymicrobial nature of the gut, an as yet unexplored theme in the context of ecological theory, is whether Blastocystis becomes pathogenic through synergistic action with other microbes and/or their metabolic products. Nonetheless, the asymptomatic carriage of Blastocystis found in our study and others reinforces the idea of it being a part of the healthy gut microbiota.
We used PCR to amplify the barcoding region of *Blastocystis* (Scicluna *et al.*, 2006) from 178 fecal samples. This method demonstrated that the overall prevalence of *Blastocystis* in our study population was 23%. It is difficult to determine how this prevalence rate compares to other studies from Thailand. *Blastocystis* detection in the country is primarily based on microscopy (Boonjaraspinyo *et al.*, 2013; Kitvatanachai and Rhongbuttsri, 2013; Kitvatanachai *et al.*, 2008; Ngrengarmlert *et al.*, 2007; Pipatsatitpong *et al.*, 2012; Sagnuankiat *et al.*, 2014). In general, microscopy is less sensitive and underestimates prevalence. Molecular-based studies on *Blastocystis* constitute a recent development and remain extremely limited (Boondit *et al.*, 2014; Jantermtor *et al.*, 2013; Palasuwan *et al.*, 2016; Pintong *et al.*, 2014; Pipatsatitpong *et al.*, 2015; Popruk *et al.*, 2015; Thathaisong *et al.*, 2013). These molecular surveys have shown prevalence ranging from 6% to 51%. The observed variable data from our study and others could be the result of demographically different populations being surveyed. For instance, the highest abundance of *Blastocystis* was found in orphan male children, while residents of communities along Chao Praya River in Ayutthaya Province (central Thailand) had the lowest prevalence (Palasuwan *et al.*, 2016; Pintong *et al.*, 2014). Age, health status of the individuals, and general lifestyle are likely contributing factors accounting for the observed differences among studies (Leelayoova *et al.*, 2008).

Subtyping of *Blastocystis* positive samples identified six of the nine STs known to colonise humans: ST1, ST2, ST3, ST4, ST6 and ST7. ST3 was dominant and present in two thirds of the samples, followed by ST1 and ST7. This is slightly different than the global trend, whereby ST3, ST1 and ST2, in this order, are the most frequently encountered (Stensvold and Clark, 2016). This pattern is also noted in molecular surveys of *Blastocystis* in neighboring Southeast Asian countries (Laos, 2013; Kitvatanachai and Rhongbuttsri, 2013; Kitvatanachai *et al.*, 2008; Ngrengarmlert *et al.*, 2007; Pipatsatitpong *et al.*, 2012; Sagnuankiat *et al.*, 2014). In general, microscopy is less sensitive and underestimates prevalence. Molecular-based studies on *Blastocystis* constitute a recent development and remain extremely limited (Boondit *et al.*, 2014; Jantermtor *et al.*, 2013; Palasuwan *et al.*, 2016; Pintong *et al.*, 2014; Pipatsatitpong *et al.*, 2015; Popruk *et al.*, 2015; Thathaisong *et al.*, 2013). These molecular surveys have shown prevalence ranging from 6% to 51%. The observed variable data from our study and others could be the result of demographically different populations being surveyed. For instance, the highest abundance of *Blastocystis* was found in orphan male children, while residents of communities along Chao Praya River in Ayutthaya Province (central Thailand) had the lowest prevalence (Palasuwan *et al.*, 2016; Pintong *et al.*, 2014). Age, health status of the individuals, and general lifestyle are likely contributing factors accounting for the observed differences among studies (Leelayoova *et al.*, 2008).

Subtyping of *Blastocystis* positive samples identified six of the nine STs known to colonise humans: ST1, ST2, ST3, ST4, ST6 and ST7. ST3 was dominant and present in two thirds of the samples, followed by ST1 and ST7. This is slightly different than the global trend, whereby ST3, ST1 and ST2, in this order, are the most frequently encountered (Stensvold and Clark, 2016). This pattern is also noted in molecular surveys of *Blastocystis* in neighboring Southeast Asian countries (Laos,
Cambodia and Malaysia), except for Laos, where ST1 was the most frequent followed by ST3 and then ST2 (Nithyamathi et al., 2016; Noradilah et al., 2017; Sanpool et al., 2017; Wang et al., 2014). Herein, ST2, ST4 and ST6 were found in a single individual each. ST4 is a geography-influenced ST and is typically found in Europe, where it is abundant and highly prevalent (Alfellani et al., 2013b; Beghini et al., 2017; Forsell et al., 2016). Indeed, ST4 occurrence in Southeast Asia is extremely rare having been found so far only in molecular surveys taking place in Thailand only two cases of ST4 have been found: one in this study and another from Tak, a northwestern province bordering Myanmar (Popruk et al., 2015). In the neighboring countries of Laos, Cambodia, and Malaysia, ST4 has been found only in Malaysia (Nithyamathi et al., 2016; Noradilah et al., 2017; Popruk et al., 2015). Regrettably, there is no publicly available data on Blastocystis from Myanmar. Alfellani et al., (2013b), noted that ST4 occurs rarely in places, where ST1 dominates, and/or in regions where Muslim populations predominate, an observation that has since been reinforced (Forsell et al., 2016). Another explanation could be that the Thai ST4 cases represent imports from Europe. Though none of the participants in our study travelled to Europe, it is unknown, if other household members did. If this indeed were the case, it would be interesting to see how the newly arrived STs affect the resident microbiota. When all available sequences from human stool samples from Thailand are taken into account (including ours), the most prevalent STs are ST3, ST1 and ST2, a pattern that follows the global trend (Stensvold and Clark, 2016). Given the known influence of geography on ST distribution, we zoomed into specific regions. Based on geography, Thailand is divided into four regions: north, northeast, center and south. Thus, when geographic region is taken into account, there are some notable differences in the ST distribution. For instance, the relative abundance of ST3 and

11
ST1 changes. ST1 is dominant in the center, while ST3 dominates in the north and northeast. ST2 is absent in the northeast, whilst ST4 has only been found so far in the north. Importantly, the center, which is the most urbanised region of Thailand, has the least number of subtypes. Incidentally, this is also the most broadly surveyed region and where most sequences come from. Thus, though we cannot exclude the possibility that some STs circulate in the center and have yet to be detected, this is not very likely. Presence of ST6 and ST7 in explaining these differences is a challenging endeavor. It is known that the four geographic regions of Thailand also happen to be gastronomically distinct. In particular, the north and northeast could be attributed to domestic fowl, and most especially chickens and geese, which live in close proximity to humans. Chiang Rai is of interest, because of its northernmost location (Fig. 1) and the mixed influence of nearby Laos and Myanmar on daily life, including diet. In a broader context, and given the influence of diet on microbial communities, the observed differences in ST distribution could also reflect differences in gut microbiota composition. In that vein, recent data have demonstrated that Blastocystis is more prevalent in individuals with increased Prevotella and Ruminococcus and decreased Bacteroides (Andersen and Stensvold, 2015; Audebert et al., 2016; Beghini et al., 2017; O’Brien Andersen et al., 2016). Nonetheless, a knowledge gap exists, as to whether similar differences drive prevalence of specific STs and vice versa.

The data obtained herein are the first on prevalence of Blastocystis and distribution of STs in asymptomatic individuals living in Chiang Rai Province, and one of the few in Thailand. Our data along with other molecular data from Thailand will contribute to the knowledge on the epidemiology of Blastocystis in the country. Nonetheless, major gaps exist, in that: 1) molecular data on Blastocystis from Thailand is scant, thus ST distribution is largely unknown and, 2) the majority of Thai
communities have yet to be sampled. For example, there is no molecular data of
Blastocystis from the south part of Thailand. Moving forward, future studies should
expand to generate subtyping information from and include more communities
across not only from Thailand and also include symptomatic individuals, but
neighboring countries as well.

CONFLICTS OF INTEREST

The authors declare no conflict of interest

ACKNOWLEDGMENTS

We are grateful to all the volunteers that participated in this study. We also wish to
acknowledge the Human Gut Microbiome for Health Research Unit, Mae Fah Luang
University, Chiang Rai, Thailand.

FINANCIAL SUPPORT

This project was supported by the Thailand Research Fund (RSA6080048) awarded to
E.G. A.D.T. was supported by BBSRC research grant (BB/M009971/1).

REFERENCES

Alfellani, M.A., Jacob, A.S., Perea, O.N., Krecek, R.C., Taner-Mulla, D., Verweij,
J.J., Levecke, B., Tannich, E., Clark, C.G. 2013a. Diversity and distribution of
Alfellani, M.A., Stensvold, C.R., Vidal-Lapiedra, A., Onuoha, E.S., Fagbenro-


Boonjaraspinyo, S., Boonmars, T., Kaewsamut, B., Ekobol, N., Laummaunwai, P.,


Blastocystis infection is associated with irritable bowel syndrome in a Mexican patient population. Parasitol Res. 110, 1269-1275.


Pintong, A., Sukthana, Y., Mori, H., Mahittikorn, A., Tungtrongchitr, R.,


Sanpool, O., Laymanivong, S., Thanchomnang, T., Rodpai, R., Sadaow, L., Phosuk,


Tan, K.S.W., Mirza, H., Teo, J.D.W., Wu, B., MacAry, P.A. 2010. Current views on


FIGURE LEGENDS

Fig. 1. Map of Thailand and Chiang Rai Province districts. Sampling localities used in this investigation are depicted in red lettering.

Fig. 2. Frequency of *Blastocystis* alleles detected in the study population

Fig. 3. Prevalence of *Blastocystis* subtypes in Thailand (n=161)

Fig. 4. Prevalence and distribution of *Blastocystis* subtypes across geographic regions in Thailand

Fig. 5. Maximum likelihood phylogenetic tree inferred from 103 SSUrRNA sequences and 1379 sites. New sequences are depicted in bold lettering. Numerical values on the tree branches indicate bootstrap support percentages and posterior probabilities in this order.