



# Genome-Wide Prediction of the Polymorphic *Ser* Gene Family in *Tetrahymena thermophila* Based on Motif Analysis

Patrath Ponsuwanna<sup>1</sup>, Krittikorn Kümpornsin<sup>1</sup>, Thanat Chookajorn<sup>2\*</sup>

<sup>1</sup> Department of Biochemistry, Faculty of Science, Mahidol University, Bangkok, Thailand, <sup>2</sup> Center of Excellence in Malaria, Faculty of Tropical Medicine, Mahidol University, Bangkok, Thailand

## Abstract

Even though antigenic variation is employed among parasitic protozoa for host immune evasion, *Tetrahymena thermophila*, a free-living ciliate, can also change its surface protein antigens. These cysteine-rich glycosylphosphatidylinositol (GPI)-linked surface proteins are encoded by a family of polymorphic *Ser* genes. Despite the availability of *T. thermophila* genome, a comprehensive analysis of the *Ser* family is limited by its high degree of polymorphism. In order to overcome this problem, a new approach was adopted by searching for *Ser* candidates with common motif sequences, namely length-specific repetitive cysteine pattern and GPI anchor site. The candidate genes were phylogenetically compared with the previously identified *Ser* genes and classified into subtypes. *Ser* candidates were often found to be located as tandem arrays of the same subtypes on several chromosomal scaffolds. Certain *Ser* candidates located in the same chromosomal arrays were transcriptionally expressed at specific *T. thermophila* developmental stages. These *Ser* candidates selected by the motif analysis approach can form the foundation for a systematic identification of the entire *Ser* gene family, which will contribute to the understanding of their function and the basis of *T. thermophila* antigenic variation.

**Citation:** Ponsuwanna P, Kümpornsin K, Chookajorn T (2014) Genome-Wide Prediction of the Polymorphic *Ser* Gene Family in *Tetrahymena thermophila* Based on Motif Analysis. PLoS ONE 9(8): e105201. doi:10.1371/journal.pone.0105201

**Editor:** Kelly A. Brayton, Washington State University, United States of America

**Received:** January 25, 2014; **Accepted:** July 21, 2014; **Published:** August 18, 2014

**Copyright:** © 2014 Ponsuwanna et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Funding:** This work was supported by the Commission of Higher Education-Thailand Research Fund-Mahidol University (RMU5380054) and the CPMO-National Science and Technology Development Agency. KK was supported by The Thailand Research Fund through the Royal Golden Jubilee PhD Program (PHD/0044/2554). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing Interests:** The authors have declared that no competing interests exist.

\* Email: thanat.cho@mahidol.edu

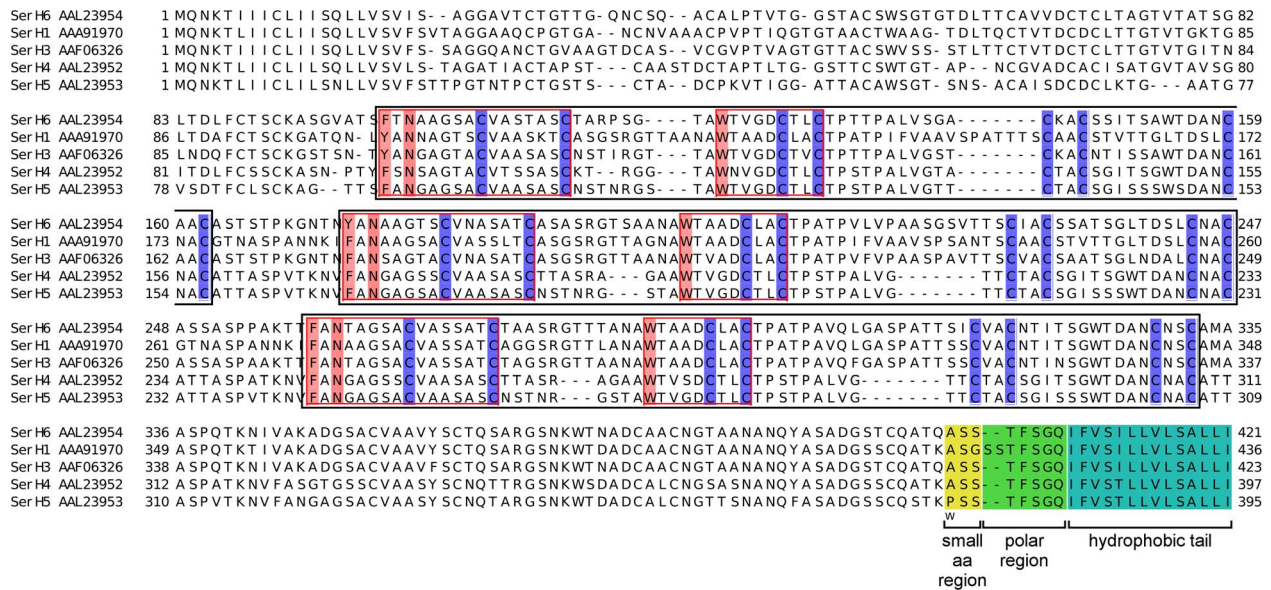
## Introduction

*Tetrahymena thermophila* is a single-celled ciliate found in temperate freshwater [1,2]. *T. thermophila* naturally feeds on bacteria, but it can also grow in media under laboratory conditions [3]. It has two nuclei, a micronucleus (MIC) as a germ line and a macronucleus (MAC) as a source specific for gene expression [4]. When food is abundant, *T. thermophila* reproduces asexually, but starvation induces conjugation between different mating types [4]. Nuclei developed from parental MICs are exchanged between two mating pair to produce new MIC and MAC. Nucleus destined to become a new MAC undergoes DNA rearrangements, including deletion of internal eliminated sequences (IES), removal of repetitive sequences and chromosome breakage at specific sites [4]. During the development of the new MAC, the parental MAC becomes degraded [4,5]. This unique biology of *T. thermophila* makes it an important model organism leading to seminal discoveries in the field of molecular biology [2].

*T. thermophila* cell membrane is covered by a surface protein known as immobilization antigen (i-ag) [6,7], as incubation with antibodies against i-ag causing *T. thermophila* to cease its movement, hence the name. Various subtypes of *T. thermophila* i-ag have been described based on immobilization assays with specific antibodies. Subtypes H, L and T are expressed at different temperatures [6,8], with subtype H being expressed at “high” temperature (20–35°C), subtype L at “low” temperature (<20°C)

[6] and subtype T (torrid) at temperatures above 36°C [8,9]. Gradual switching of i-ag subtypes occurs when temperatures are shifted [6,8]. The gene coding for i-ag was named *Ser* after the word “serotype” [10]. So far, up to six subtype H allelic variants, one subtype J gene and six subtype L paralogs were found [11–15]. One common characteristic among these *Ser* proteins is a repetitive cysteine-rich motif [13,14]. Such features are also common in surface proteins of other unicellular eukaryotes [16]. The control mechanism of *Ser* expression is not well understood, but the mRNA half-life of *SerH3* (normally expressed at 20–35°C) is rapidly decreased when temperature is shifted up to 40°C [17,18]. Treatment with protein synthesis and protein kinase inhibitors can prolong *SerH3* mRNA half-life during this temperature shift, suggesting that there are proteins and phospho-proteins involved in *Ser* mRNA degradation [19]. However, the role of *T. thermophila* i-ag remains unclear, though it may involve sensing the environment or prey-predator recognition, similar to *Paramecium* surface antigen [20].

Even though sequence variation is a hallmark of these highly diverged surface proteins, they often contain repetitive cysteine-rich motifs. The periodic cysteine residues could form disulfide bonds in a consistent pattern among proteins in the same family, but there is no experimental data on the formation of disulfide bonds in i-ag at this point. The existence of the disulfide bonds might introduce an extremely hydrophobic moiety at the core of



**Figure 1. Features of *T. thermophila* Ser protein sequences.** Selected known Ser protein sequences are aligned to show their Cys pair, repetitive block and GPI anchor site. Cys residues are highlighted (dark blue). Each repetitive block is indicated by black box. Red box indicates sequence feature which appears in each repetitive block. The region predicted as GPI anchor signal by FragAnchor is color-shaded. Predicted GPI attachment site is marked by letter “w”. According to data from known Ser, number of Cys residues per repetitive block is unique for each subtype (SerL: 6 Cys per block; SerH: 8 Cys per block). The length of each repetitive block differs among various subtypes and is also varied between 55–100 aa. GPI anchor signal predicted by FragAnchor exhibits region of small amino acids (Ala, Gly, Ser) where GPI is attached (yellow), followed by polar region (green) and hydrophobic tail (light blue). doi:10.1371/journal.pone.0105201.g001

protein [21]. It was suggested that ‘hydrophobic collapse’ might play a crucial role in protein folding in general via hydrophobic core nucleus which drives the folding process [22]. Disulfide bond formation could allow protein to become highly divergent on the surface while maintaining the overall fold. Different number of cysteine pair per one repeating sequence block in different i-ag subtypes has been documented [12]. Despite high sequence variation among Ser genes, the pattern of cysteine rich motif is a common feature among them.

Another distinct feature among Ser genes is the consensus sequence at the C-terminus specific for glycosylphosphatidylinositol (GPI) anchor modification. *T. thermophila* i-ag subtype H was shown to be GPI-anchored protein by radiolabelling [23,24]. Putative GPI anchor site was predicted to be located at the C-terminus of the Ser proteins [13,14]. There is a GPI anchor signal sequence at the C-terminus, which can be recognized by transamidase. GPI anchor signal can be divided into three regions: GPI attachment site ( $\omega$  site), spacer region of polar residues ( $\omega+3$  to  $\omega+8$ ), and hydrophobic region ( $\omega+9$  to the C-terminus) [25]. There may be the minimum length for hydrophobic region required for GPI attachment [26]. Because GPI signal has no detectable conserved sequence, it cannot be identified by sequence similarity approach. However, the signal sequences surrounding the GPI attachment site can be defined as regions of amino acid residues with different physical properties such as size and hydrophobicity [25]. This allows the prediction of GPI-anchored protein using knowledge-based algorithm [27].

There are 24,725 predicted protein-coding genes in *T. thermophila* MAC genome [28,29]. The majority of the genes in *T. thermophila* are transcriptionally regulated as determined by nuclear run-on assay [30]. Post-transcriptional control via mRNA stability was also observed in SerH3 gene regulation [18]. As the Ser gene family is highly polymorphic, sequence homology analysis alone cannot recognize the full set of putative Ser genes.

In order to gain a full understanding of the antigenic variation in *T. thermophila* Ser genes, it is necessary to overcome this problem. In this study, the repertoire of the *T. thermophila* Ser gene family was annotated by setting search criteria based on the repetitive cysteine-rich motif and the signal sequence for GPI anchor. Two hundred and sixteen putative Ser genes from *T. thermophila* MAC genome sequence were selected including the known Ser genes. Previous studies have shown that each Ser subtype has a specific number of Cys residues per repetitive block: SerH with 8 Cys residues per block [14,15], SerJ with 10 Cys residues per block [12] and SerL with 6 Cys residues per block. The periodic cysteine block pattern,  $CX_{(\geq 6)}CX_{(\geq 1)}CX_{(\geq 1)}CX_{(\geq 1)}CX_{(\geq 1)}C$ , and the GPI anchor site were used as the search criteria for the Ser gene family (Figure 1 and Figure S1 in File S1) [13–15]. Putative Ser genes were further characterized based on their phylogenetic distribution with experimentally classified Ser subtypes. By combining the data from the Ser candidates and the known Ser genes, the patterns of Ser chromosomal localization and gene expression were revealed.

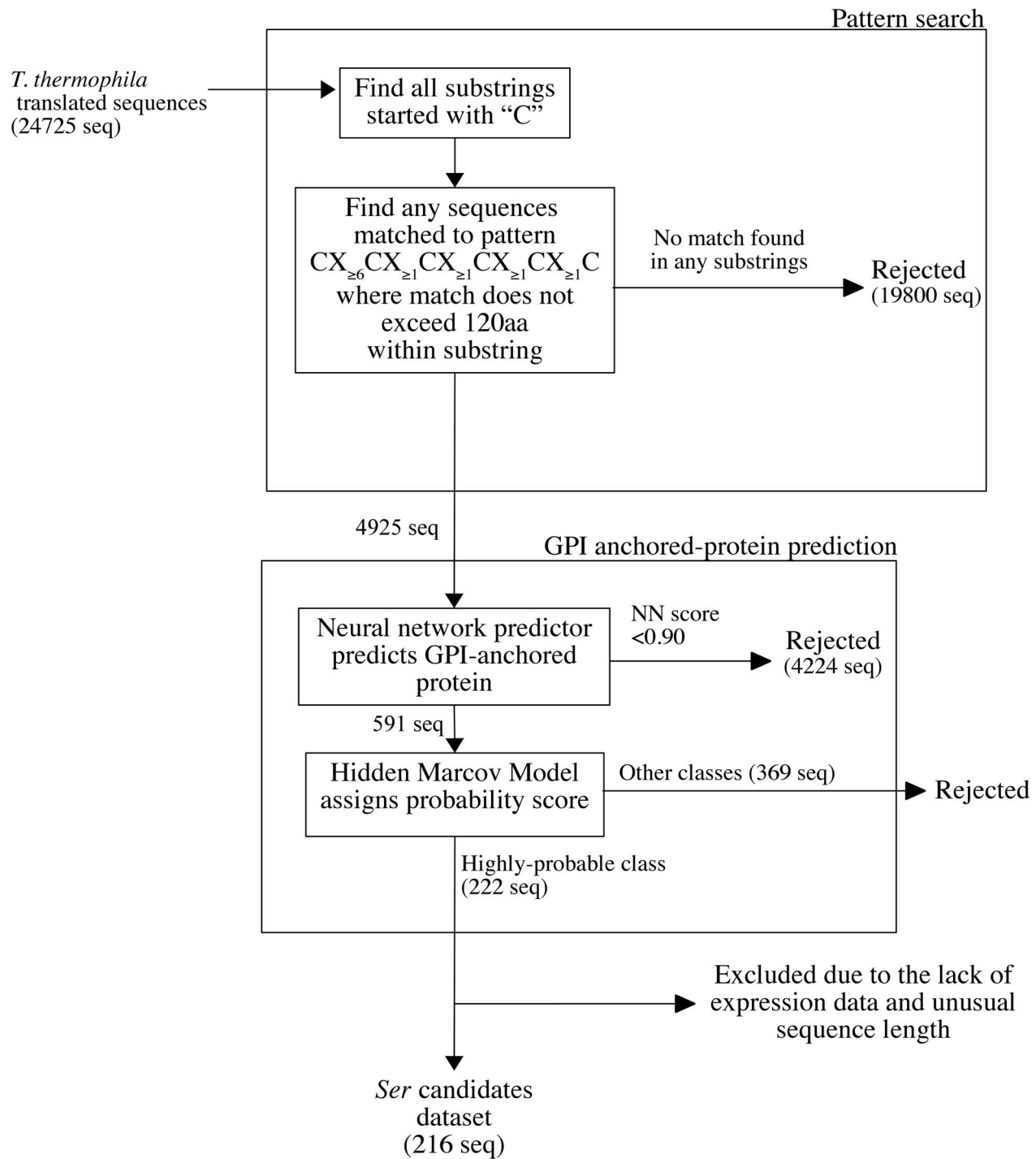
**Materials and Methods**

**Sequence and genome data**

The whole predicted protein sequences of *Tetrahymena thermophila* were downloaded from *Tetrahymena* Genome Database (<http://ciliate.org/index.php/home/downloads>). The 5'-UTR sequences of Ser candidates were obtained from TetraFGD Genome Browser (<http://tfgdgb.ihb.ac.cn>). Thirteen known Ser subtypes (listed in Table S1 in File S1) from NCBI database were used in this study as reference [28,29].

**Identification of Ser gene candidates**

Ser prediction algorithm was composed of two main parts, cysteine pattern search and GPI-anchored protein prediction (Figure 2). Custom Perl script was used to perform the search for the cysteine-

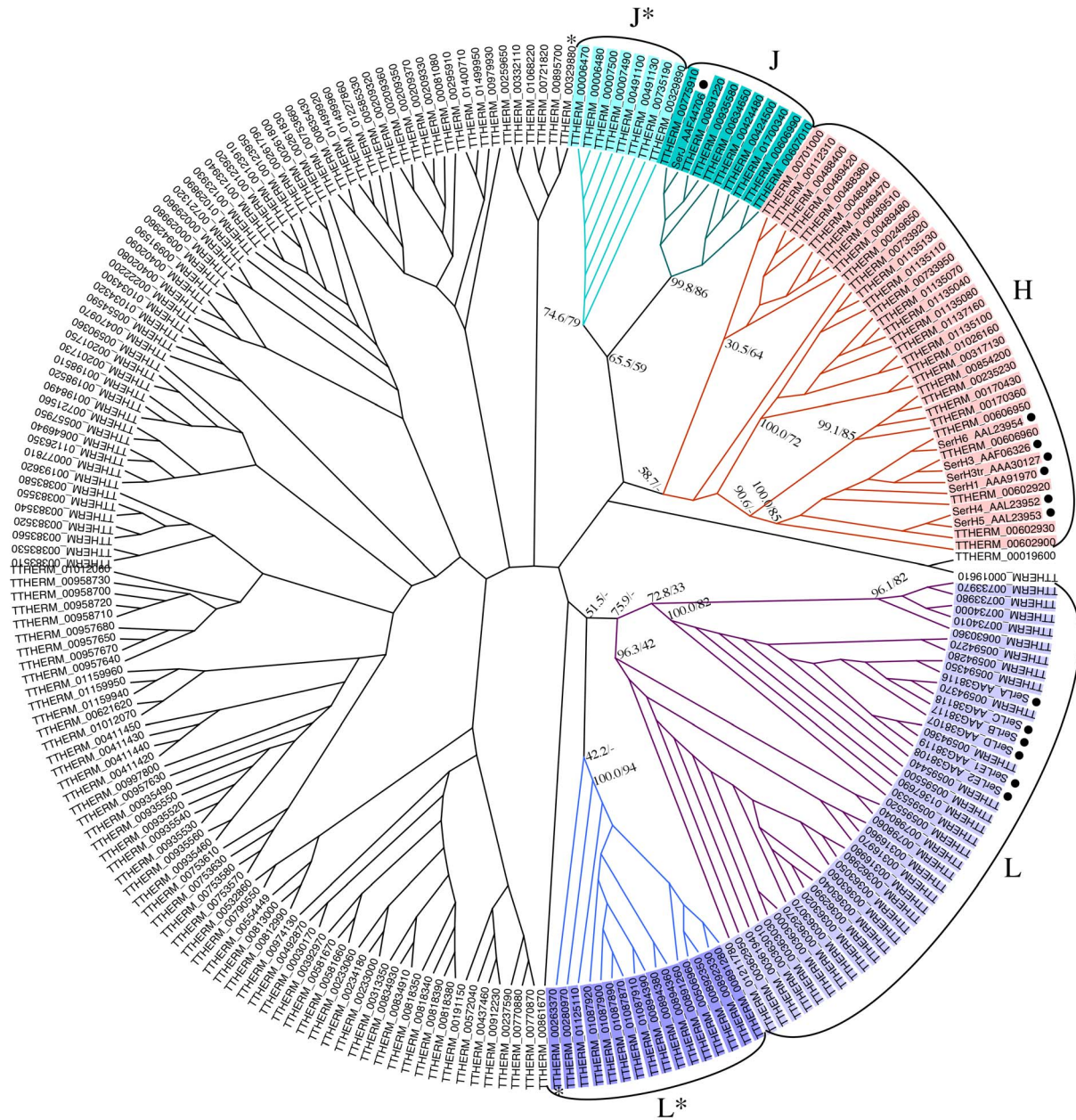


**Figure 2. *T. thermophila* Ser prediction algorithm diagram.** *T. thermophila* translated sequences were the input data for the pattern search script. *Ser* candidate detection pattern is based on length and number of Cys in a single repetitive block from known *Ser* data. In order to find all possible pattern matches within one protein sequence, the script breaks the input sequence into smaller substrings starting with C (Cys). Then the script finds the match within that substring. *T. thermophila* translated sequences which pass pattern search algorithm were then input into web-based GPI-anchored protein predictor FragAnchor. FragAnchor assigned predicted GPI anchored proteins into classes based on their HMM scores. Sequences assigned into the “highly probable” class were selected as *Ser* candidates. doi:10.1371/journal.pone.0105201.g002

rich pattern. The script was set to select any sequence containing  $CX_{\geq 6}CX_{\geq 1}CX_{\geq 1}CX_{\geq 1}CX_{\geq 1}C$  or  $CX_{\text{long}}CX_{\text{short}}C$  (C = cysteine, X = any amino acid except cysteine,  $X_{\text{short}} = 1$  to 5 residues and  $X_{\text{long}} > 5$  residues) within 120 amino acid residues. This

length of amino acid residues was chosen because it gave the best pattern match to the known *Ser* genes. The results were then refined by GPI-anchored protein prediction employing the web-based program FragAnchor [31]. Hidden Markov Model (HMM) imple-





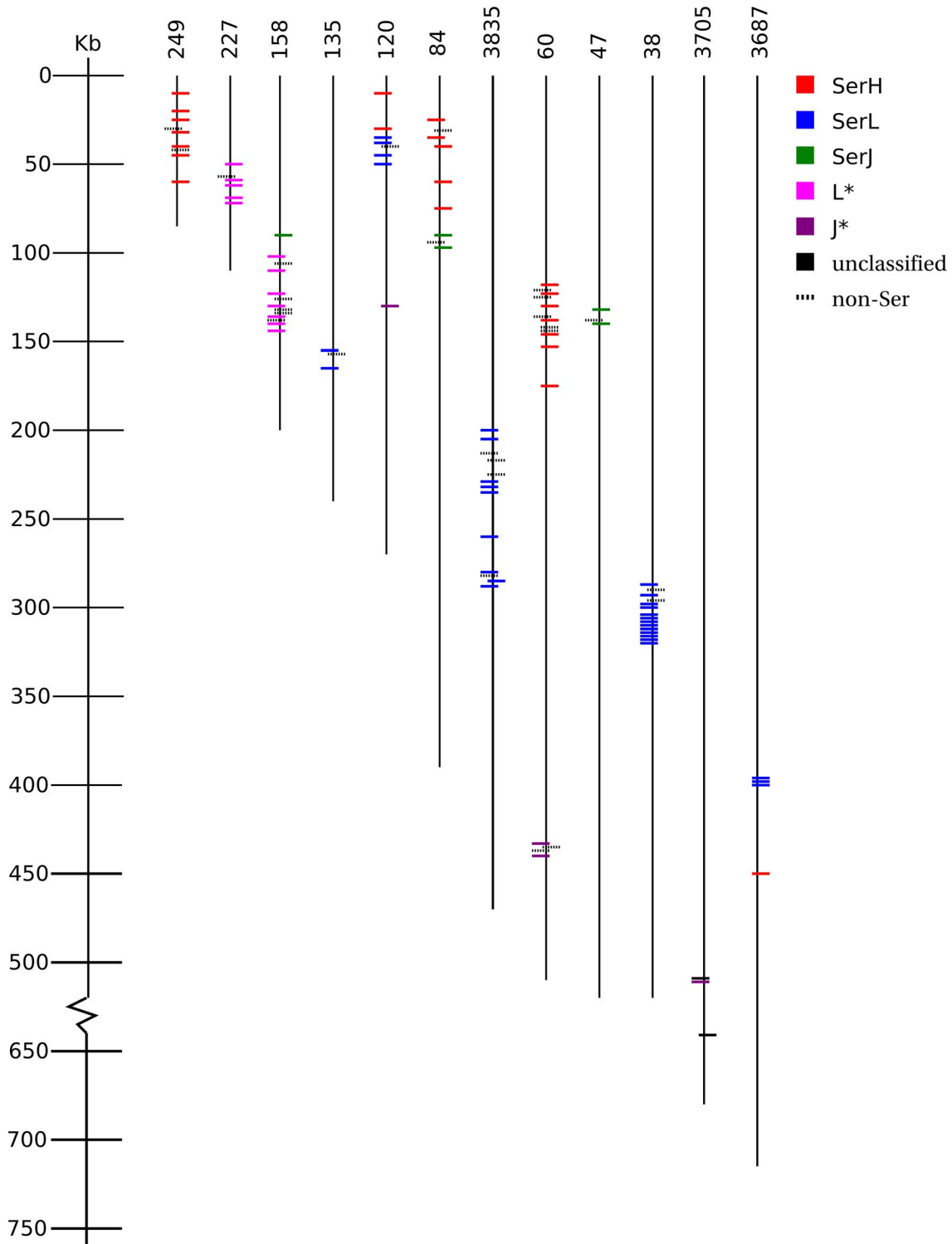
**Figure 3. Neighbor-joining cladogram of *Ser* candidates identified in this study.** Known *Ser* proteins are also included in this figure (black dots). 100-replicate bootstrapping was performed. Group-assigned candidates are highlighted. Bootstrap support percentages for the Neighbor-Joining (NJ) tree and the Maximum Likelihood (ML) tree (dash indicates undetermined bootstrap support value) are respectively shown on selected branch nodes. Among group-assigned candidates, TtHERM\_00263370 and TtHERM\_00329880 are inconsistent between NJ and ML trees. In NJ tree, TtHERM\_00263370 is clustered with L\* clade. But in ML tree, it is closer to SerL clade. TtHERM\_00329880 is clustered with SerH clade in the ML tree but not in the NJ tree. doi:10.1371/journal.pone.0105201.g003

mented in FragAnchor was used to assign the probability score. Amino acid sequences matching the defined cysteine pattern with highly probable GPI-anchored protein determined by FragAnchor were selected as *Ser* candidates.

**Phylogenetic analysis**

The translated sequences of *Ser* candidates were aligned using ClustalX 2.0.12 with default multiple alignment parameters (gap opening penalty = 10; gap extension penalty = 0.2; Gonnet series weight matrix). Alignment was then adjusted manually. Neighbor-

joining (NJ) tree was calculated with 1000 bootstrap replicates using ClustalX. Protein evolutionary model was selected by ProtTest [32]. Maximum likelihood (ML) tree was estimated using RAxML 7.2.8 as implemented on the CIPRES Science Gateway [33–35]. Phylogenetic trees were then created using Dendroscope [36]. TtHERM\_01098980 was excluded from phylogenetic analysis because its sequence has an unusual sequence length of 3751 amino acid residues, preventing it from being aligned.



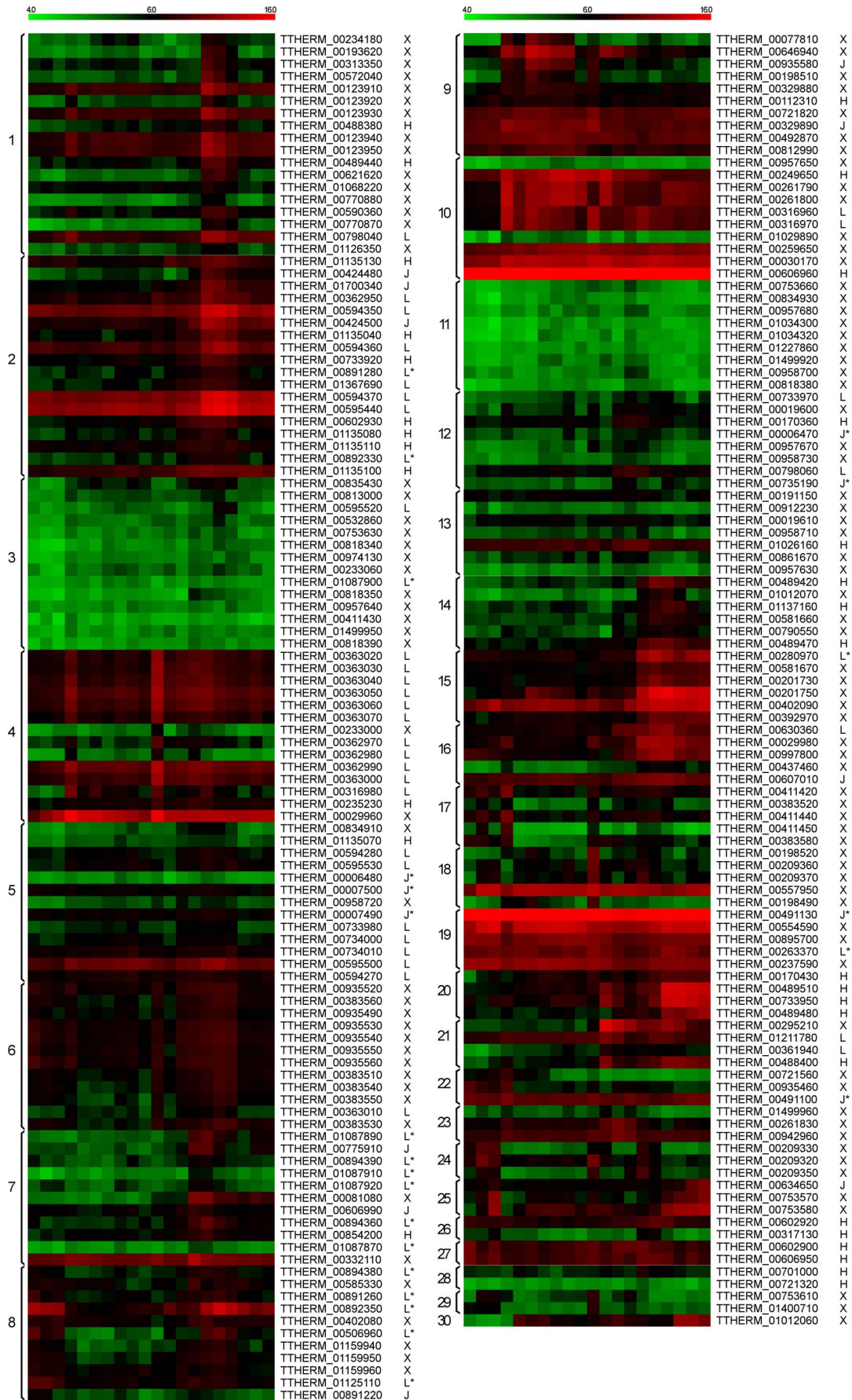
**Figure 4. Distribution of *Ser* candidates on *T. thermophila* MAC scaffold.** Each short horizontal line represents one *Ser* gene. *Ser* subtype is color-coded. Dash line represents non-*Ser* gene which locates within *Ser* tandem arrays. Gene orientation on plus or minus strand is depicted by left or right alignment respectively. Scaffold number is shown on the top of each scaffold line (vertical line). Only scaffolds containing classified *Ser* tandem array are shown.

doi:10.1371/journal.pone.0105201.g004

***Ser* gene expression analysis**

*T. thermophila* genome-wide gene expression (growth, starvation and conjugation) microarray data was retrieved from

*Tetrahymena* Functional Genomics Database (TetraFGD) (<http://tfgd.ihb.ac.cn/>) [37]. MultiExperimentViewer (MeV) was employed to evaluate gene expression clusters and to classify the





**Figure 5. *Ser* gene expression clusters.** Each row represents one gene. Gene ID and subtype were listed. Unclassified *Ser* candidates are marked as X. Gene expression data was subjected to K-means clustering method using Pearson correlation to measure distance. Expression cluster ID is indicated as arabic number on the left of expression heatmap. Scale bar represents log<sub>2</sub>-transformed gene expression value. Red indicates expression value above median, and green indicates expression value below median. Clustering analysis was performed using MeV 4.7.  
doi:10.1371/journal.pone.0105201.g005

expression pattern into subgroups (www.tm4.org/mev/). Clustering was identified with K-Means clustering module (KMC), and distance was calculated by Pearson correlation.

**Results**

**Identification and classification of *Ser* genes**

Due to the high degree of polymorphism, identification of *T. thermophila Ser* genes by sequence homology alone is limited by low sequence conservation. In order to systematically search for *Ser* candidates, two criteria were applied based on the common features found in existing i-ag proteins (six *SerH*, one *SerJ* and six *SerL*) namely, the presence of Cys residue pattern block CXlongCXshortC and the GPI-anchor signal located at the C-terminus. For the first criterion, a Perl script was set up to search for the *T. thermophila* proteins which, for any window frame containing 4 or 6 Cys residues within 30–120 amino acid residues, the number of amino acid between the first Cys pair is more than 5 and the number of amino acid between the other Cys pairs is at least 1 (Figure 2). After determining the number of hits versus search pattern for saturation in gene numbers and change in phylogenetic pattern, the search criterion was limited to the minimum of 6 Cys residues within a 120 amino acid interval. Any protein containing CX(≥6)CX(≥1)CX(≥1)CX(≥1)CX(≥1)C sequence within 120 amino acid residues was selected, resulting in 4,925 hits out of 24,725 *T. thermophila* predicted proteins. For the second selection criterion, FragAnchor classified 216 *Ser* candidates as “highly probable” GPI-anchored proteins including all known *Ser* genes (Figure 1–2 and Figure S1 in File S1). Sequences classified

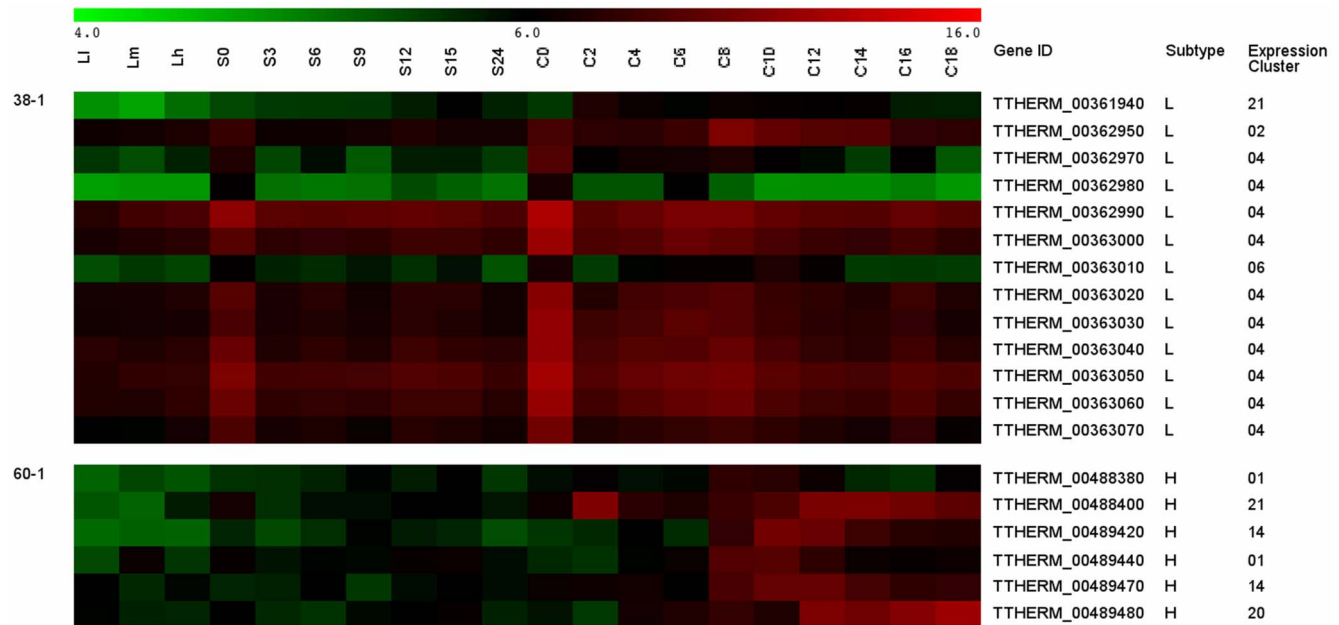
as “probable” or “weakly-probable” GPI-anchored proteins were excluded in order to avoid false positives. The approach successfully identified all four annotated *Ser* genes in the genome of *T. thermophila* strain SB210 with perfect identity match to the experimentally identified *Ser* genes.

**Classification of putative *Ser* genes from *T. thermophila* strain SB210**

Forty-five percent of the selected genes could be grouped with three identified *Ser* subtypes, *SerH*, *SerL* and *SerJ*, with good bootstrap support from NJ analyses (*SerH*: 58.7% NJ bootstrap support; *SerL*: 72.8% NJ bootstrap support; and *SerJ*: 99.8% NJ bootstrap support). Two distinct branches of *Ser* candidates, grouped with *SerJ* and *SerL*, were found and were named J\* (74.6% NJ bootstrap support) and L\* (42.2% NJ bootstrap support) to reflect their phylogenetic association with *SerJ* and *SerL* groups, respectively (Figure 3). Candidates grouped with the known *Ser* subtypes are listed in Table S2 in File S1.

***Ser* chromosomal location**

Chromosomal location analysis focused on the newly identified *Ser* genes with known subtypes. They were dispersed on 76 out of the estimated 250–300 MAC chromosomes in *T. thermophila*. Figure 4 shows the distribution of subtype-classified *Ser* candidates on 12 scaffolds, with the majority (86%) of the identified *Ser* candidate genes located in close proximity to one another. *Ser* genes tend to form a tandem array of the same subtype (Table S2 in File S1). Six MAC scaffolds contained tandem arrays with only



**Figure 6. Expression profile of *Ser* genes from tandem array 38-1 and 60-1.** Gene expression data was collected from *T. thermophila* during growth at low (LI), medium (Lm), and high (Lh) densities. The data from starvation (S0, S3, S6, S9, S12, S15, S24) and conjugation (C0, C2, C4, C6, C8, C10, C12, C14, C16, C18) samples was also included with numerical values showing hours in each particular condition [37]. Complete *Ser* expression cluster result is available in Table S2 and S3 in File S1.  
doi:10.1371/journal.pone.0105201.g006

one subtype of *Ser* genes, and those in close proximity on the same scaffold tended to have the same orientation, with a few exceptions, such as subtype-H TTHERM\_00602920 on scaffold 84 and subtype-L TTHERM\_00595520 on scaffold 3835. There appears to be no preference for scaffold size or chromosomal region where *Ser* genes are located.

### Ser gene expression analysis

*Ser* expression patterns during growth, starvation and conjugation were analyzed based on DNA microarray data [37]. *Ser* expression data could be grouped into 30 clusters based on their expression patterns (Figure 5). No correlation was found between specific *Ser* subtypes and expression patterns. Certain *Ser* tandem genes were associated with the same expression pattern. For example, ten out of thirteen subtype-L *Ser* genes on scaffold 38 were found in expression cluster 4 (Figure 6, upper panel). However, some tandem arrays were not in the same expression cluster, but their expression patterns appeared to be stage-specific. For example, scaffold 60 contains a tandem array of 6 subtype-H genes grouped into four different expression clusters (Figure 6, lower panel). Interestingly, they were all up-regulated during conjugation, but at different time points. Data of similar analyses on locations and expression patterns of unclassified *Ser* candidates are available in Table S3 in File S1.

### Discussion

*T. thermophila* i-ag was originally identified as variant surface antigen. Their high degree of polymorphism has prevented systematic identification of its *Ser* family based on sequence homology alone despite available genomic data. In this study, an alternative strategy was adopted for identifying this gene family by selecting two common features of the known *Ser* genes, namely, the Cys-rich motif and GPI-anchor site. Their phylogenetic distributions were analyzed in order to characterize *Ser* candidates that are related to the known subtypes. The approach identified *Ser* candidates that can be grouped into the known subtypes. However, 55% of the genes could not be classified into any known subtype based on phylogenetic analysis, and might belong to new subtypes. At present, the genes encoding several types of surface antigens such as *SerT*, *SerS*, *SerM* and *SerI* have not been identified indicating the existence of more *Ser* repertoires and subtypes [38]. In addition, certain unclassified gene candidates exhibit properties similar to *Ser*, such as tendency to be located in tandem and have similar gene expression profile. It is possible that the missing *Ser* subtypes might belong to one of the unclassified gene families. Experiments with specific antibodies to *Ser* candidates will be required in order to prove that they are indeed i-ag proteins. Sequence comparison and synteny analysis of the highly polymorphic *Ser* genes from more *T. thermophila* isolates will confirm whether these genes are under positive selection which is a strong evolutionary driving force for mating proteins, molecular sensors and evasive decoys [39,40]. Using the *Ser* candidates as blastp query on NCBI protein database yields sixteen *T. thermophila* proteins not previously included in the list [Table S4 in File S1]. Nevertheless, they either lack the GPI anchor motif or matching cysteine pattern. The phylogenetic analysis showed that these proteins are not grouped with any *Ser* candidate. The analysis outside *T. thermophila* revealed two weak

hits with the proteins from other ciliates [*Ichthyophthirius multifiliis* AAK94941 with 21% identity to *SerL* and *Paramecium tetraurelia* XP\_001450224 with 22% identity to *SerH*]. This might indicate that the *Ser* gene family is unique to *T. thermophila*.

i-ag proteins were discovered based on their variation in response to antibodies directed at *T. thermophila* surface antigens. This is a hallmark for many antigenic variation phenomena among parasitic and free-living protozoa [41]. In parasitic protozoa such as *Plasmodium falciparum*, a family of proteins on infected red blood cells is needed for a parasitic adherence mechanism to human cells and tissue which is crucial for malaria pathogenesis [42]. Antigenic variation in *P. falciparum* switches the variant pathogenic proteins in order to avoid immune detection [43]. Other parasitic protozoa also exploit a similar system [44,45]. When antigenic variation is compromised, the parasite becomes vulnerable to the immune system [46].

Our study took advantage of the available *T. thermophila* MAC genome and DNA microarray expression data to analyze the *Ser* gene family. These *Ser* genes are organized in tandem arrays on several MAC scaffolds. These tandem arrays often belong to the same subtype, suggesting that they arose by gene duplication or genetic recombination. Expression pattern cluster analysis does not explicitly indicate the role of the *Ser* gene family in any developmental stage in particular. Available microarray data also did not include every known culture conditions for inducing the expression of *Ser* genes. However, stage-specific expression patterns of several *Ser* transcripts at the same time point were observed, and thus mutual exclusion mechanism is not likely the only strategy underlying expression control of every *Ser* gene.

Unlike parasitic protozoa, free-living protozoa are not subject to host immune pressure, and the purpose for having surface antigenic variation remains unclear. Nevertheless, understanding the mechanism of antigenic variation in free-living organisms could provide new insights into the evolution and regulation control of antigenic variation in parasitic organisms. Thus identifying the whole repertoire of the *Ser* gene family is the first step towards the exploration of antigenic variation in *T. thermophila*, an important model organism for many seminal discoveries in molecular biology.

### Supporting Information

**File S1 Supporting information.**  
(DOC)

### Acknowledgments

The authors acknowledge P. Wilairat (Mahidol University, Thailand) for critical reading of the manuscript and N. Karoonuthaisiri (National Science and Technology Development Agency, Thailand) for assistance with DNA microarray data analysis. This work was supported by the Commission of Higher Education-Thailand Research Fund-Mahidol University (RMU5380054) and CPMO-National Science and Technology Development Agency. KK was supported by The Thailand Research Fund through the Royal Golden Jubilee PhD Program (PHD/0044/2554)

### Author Contributions

Conceived and designed the experiments: PP TC. Performed the experiments: PP. Analyzed the data: PP KK TC. Contributed reagents/materials/analysis tools: PP. Wrote the paper: PP KK TC.

### References

- Collins K, Gorovsky MA (2005) *Tetrahymena thermophila*. *Curr Biol* 15: R317–318.
- Simon EM, Nannoy DL, Doerder FP (2007) The “*Tetrahymena pyriformis*” complex of cryptic species. *Biodiversity and Conservation* 17: 365–380.



3. Elliott AM (1959) Biology of Tetrahymena. Annual Review of Microbiology 13: 79–96.
4. Karrer KM (2000) Tetrahymena genetics: two nuclei are better than one. Methods Cell Biol 62: 127–186.
5. Harrison GS, Karrer KM (1985) DNA synthesis, methylation and degradation during conjugation in Tetrahymena thermophila. Nucleic Acids Res 13: 73–87.
6. Margolin P, Loefer JB, Owen RD (1959) Immobilizing Antigens of Tetrahymena pyriformis. Journal of Eukaryotic Microbiology 6: 207–215.
7. Williams NE, Doerder FP, Ron A (1985) Expression of a cell surface immobilization antigen during serotype transformation in Tetrahymena thermophila. Mol Cell Biol 5: 1925–1932.
8. Phillips RB (1967) Inheritance of T serotypes in Tetrahymena. Genetics 56: 667–681.
9. Smith DL, Doerder FP (1992) Exceptions to mutual exclusion among cell surface i-antigens of Tetrahymena thermophila during salt stress and stationary phase. J Protozool 39: 628–635.
10. Doerder FP, Berkowitz MS, Skalican-Crowe J (1985) Isolation and genetic analysis of mutations at the SerH immobilization antigen locus of Tetrahymena thermophila. Genetics 111: 273–286.
11. Tondravi MM, Willis RL, Love HD, Jr., Bannon GA (1990) Molecular characterization of SerH3, a Tetrahymena thermophila gene encoding a temperature-regulated surface antigen. Mol Cell Biol 10: 6091–6096.
12. Doerder FP (2000) Sequence and expression of the SerJ immobilization antigen gene of Tetrahymena thermophila regulated by dominant epistasis. Gene 257: 319–326.
13. Doerder FP, Gerber CA (2000) Molecular characterization of the SerL paralogs of Tetrahymena thermophila. Biochem Biophys Res Commun 278: 621–626.
14. Gerber CA, Lopez AB, Shook SJ, Doerder FP (2002) Polymorphism and selection at the SerH immobilization antigen locus in natural populations of Tetrahymena thermophila. Genetics 160: 1469–1479.
15. Deak JC, Doerder FP (1995) Sequence, codon usage and cysteine periodicity of the SerH1 gene and in the encoded surface protein of Tetrahymena thermophila. Gene 164: 163–166.
16. Kusch J, Schmidt HJ (2001) Genetically controlled expression of surface variant antigens in free-living protozoa. J Membr Biol 180: 101–109.
17. Hallberg RL, Kraus KW, Findly RC (1984) Starved Tetrahymena thermophila cells that are unable to mount an effective heat shock response selectively degrade their rRNA. Mol Cell Biol 4: 2170–2179.
18. Love HD, Jr., Allen-Nash A, Zhao QA, Bannon GA (1988) mRNA stability plays a major role in regulating the temperature-specific expression of a Tetrahymena thermophila surface protein. Mol Cell Biol 8: 427–432.
19. McMillan PJ, Stanley JS, Bannon GA (1995) Evidence for the requirement of protein synthesis and protein kinase activity in the temperature regulated stability of a Tetrahymena surface protein mRNA. Nucleic Acids Res 23: 942–948.
20. Simon MC, Schmidt HJ (2007) Antigenic variation in ciliates: antigen structure, function, expression. J Eukaryot Microbiol 54: 1–7.
21. Chookajorn T, Kachroo A, Ripoll DR, Clark AG, Nasrallah JB (2004) Specificity determinants and diversification of the Brassica self-incompatibility pollen ligand. Proc Natl Acad Sci U S A 101: 911–917.
22. Chang JY (2011) Diverse pathways of oxidative folding of disulfide proteins: underlying causes and folding models. Biochemistry 50: 3414–3431.
23. Ko YG, Thompson GA, Jr. (1992) Immobilization antigens from Tetrahymena thermophila are glycosyl-phosphatidylinositol-linked proteins. J Protozool 39: 719–723.
24. Ron A, Williams NE, Doerder FP (1992) The immobilization antigens of Tetrahymena thermophila are glycoproteins. J Protozool 39: 508–510.
25. Eisenhaber B, Bork P, Eisenhaber F (1998) Sequence properties of GPI-anchored proteins near the omega-site: constraints for the polypeptide binding site of the putative transamidase. Protein Eng 11: 1155–1161.
26. Udenfriend S, Kodukula K (1995) How glycosylphosphatidylinositol-anchored membrane proteins are made. Annu Rev Biochem 64: 563–591.
27. Eisenhaber B, Bork P, Eisenhaber F (1999) Prediction of potential GPI-modification sites in proprotein sequences. J Mol Biol 292: 741–758.
28. Eisen JA, Coyne RS, Wu M, Wu D, Thiagarajan M, et al. (2006) Macronuclear genome sequence of the ciliate Tetrahymena thermophila, a model eukaryote. PLoS Biol 4: e286.
29. Coyne RS, Thiagarajan M, Jones KM, Wortman JR, Tallon LJ, et al. (2008) Refined annotation and assembly of the Tetrahymena thermophila genome sequence through EST analysis, comparative genomic hybridization, and targeted gap closure. BMC Genomics 9: 562.
30. Stargell LA, Karrer KM, Gorovsky MA (1990) Transcriptional regulation of gene expression in Tetrahymena thermophila. Nucleic Acids Res 18: 6637–6639.
31. Poisson G, Chauve C, Chen X, Bergeron A (2007) FragAnchor: a large-scale predictor of glycosylphosphatidylinositol anchors in eukaryote protein sequences by qualitative scoring. Genomics Proteomics Bioinformatics 5: 121–130.
32. Abascal F, Zardoya R, Posada D (2005) ProtTest: selection of best-fit models of protein evolution. Bioinformatics 21: 2104–2105.
33. Stamatakis A (2006) RAXML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. Bioinformatics 22: 2688–2690.
34. Stamatakis A, Hoover P, Rougemont J (2008) A rapid bootstrap algorithm for the RAXML Web servers. Syst Biol 57: 758–771.
35. Miller MA, Pfeiffer W, Schwartz T (2010) Creating the CIPRES Science Gateway for inference of large phylogenetic trees. Gateway Computing Environments Workshop (GCE), 2010: 1–8.
36. Huson DH, Richter DC, Rausch C, Dezulian T, Franz M, et al. (2007) Dendroscope: An interactive viewer for large phylogenetic trees. BMC Bioinformatics 8: 460.
37. Miao W, Xiong J, Bowen J, Wang W, Liu Y, et al. (2009) Microarray analyses of gene expression during the Tetrahymena thermophila life cycle. PLoS One 4: e4429.
38. Smith DL, Berkowitz MS, Potoczak D, Krause M, Raab C, et al. (1992) Characterization of the T, L, I, S, M and P cell surface (immobilization) antigens of Tetrahymena thermophila: molecular weights, isoforms, and cross-reactivity of antisera. J Protozool 39: 420–428.
39. Swanson WJ (2003) Adaptive evolution of genes and gene families. Curr Opin Genet Dev 13: 617–622.
40. Chookajorn T, Costanzo MS, Hartl DL, Deitsch KW (2007) Malaria: a peek at the var variorum. Trends Parasitol 23: 563–565.
41. Templeton TJ (2007) Whole-genome natural histories of apicomplexan surface proteins. Trends Parasitol 23: 205–212.
42. Miller LH, Ackerman HC, Su XZ, Wellem TE (2013) Malaria biology and disease pathogenesis: insights for new treatments. Nat Med 19: 156–167.
43. Chookajorn T, Ponsuwanna P, Cui L (2008) Mutually exclusive var gene expression in the malaria parasite: multiple layers of regulation. Trends Parasitol 24: 455–461.
44. Taylor JE, Rudenko G (2006) Switching trypanosome coats: what's in the wardrobe? Trends Genet 22: 614–620.
45. Lujan HD (2011) Mechanisms of adaptation in the intestinal parasite Giardia lamblia. Essays Biochem 51: 177–191.
46. Rivero FD, Saura A, Prucca CG, Carranza PG, Torri A, et al. (2010) Disruption of antigenic variation is crucial for effective parasite vaccine. Nat Med 16: 551–557, 551p following 557.