Published online 4 November 2010 Nucleic Acids Research, 2011, Vol. 39, Database issue **D625–D631**
doi:10.1093/nar/gkq1111

Full-parasites: database of full-length cDNAs of apicomplexa parasites, 2010 update

Josef Tuda[1], Arthur E. Mongan[1], Mohammed E. M. Tolba[2,3], Mihoko Imada[2,4], Junya Yamagishi[5], Xuenan Xuan[5], Hiroyuki Wakaguri[2], Sumio Sugano[2], Chihiro Sugimoto[6] and Yutaka Suzuki[2,*]

[1]Faculty of Medicine, Sam Ratulangi University, Kampus Unsrat, Bahu Manado, 95115, Indonesia, [2]Department of Medical Genome Sciences, Graduate School of Frontier Sciences, The University of Tokyo, 5-1-5 Kashiwanoha, Kashiwa, Chiba 277-8562, Japan, [3]Department of Parasitology, Assiut University, Assiut, 71515, Egypt, [4]Department of Tropical Medicine and Parasitology School of Medicine, Keio University, 35 Shinanomachi Sinjuku, Tokyo 160-8582, [5]National Research Center for Protozoan Diseases, Obihiro University of Agriculture and Veterinary Medicine, Inada-cho west 2-13, Obihiro, Hokkaido 080-8555 and [6]Research Center for Zoonosis Control, Hokkaido University, North 20, West 10 Kita-ku, Sapporo 001-0020, Japan

Received September 15, 2010; Revised and Accepted October 18, 2010

ABSTRACT

Full-Parasites (http://fullmal.hgc.jp/) is a transcriptome database of apicomplexa parasites, which include *Plasmodium* and *Toxoplasma* species. The latest version of Full-Parasites contains a total of 105 786 EST sequences from 12 parasites, of which 5925 full-length cDNAs have been completely sequenced. Full-Parasites also contain more than 30 million transcription start sites (TSS) for *Plasmodium falciparum* (Pf) and *Toxoplasma gondii* (Tg), which were identified using our novel oligo-capping-based protocol. Various types of cDNA data resources were interconnected with our original database functionalities. Specifically, in this update, we have included two unique RNA-Seq data sets consisting of 730 million mapped RNA-Seq tags. One is a dataset of 16 time-lapse experiments of cultured bradyzoite differentiation for Tg. The other dataset includes 31 clinical samples of Pf. Parasite RNA was extracted together with host human RNA, and the extracted mixed RNA was used for RNA sequencing, with the expectation that gene expression information from the host and parasite would be simultaneously represented. By providing the largest unique full-length cDNA and dynamic transcriptome data, Full-Parasites is useful for understanding host–parasite interactions and will help to eventually elucidate how monophyletic organisms have evolved to become parasites by adopting complex life cycles.

INTRODUCTION

Parasites in the phylum Apicomplexa, which includes *Plasmodium* (malaria parasites) and *Toxoplasma* species, cause worldwide health problems that require immediate action. Because genome sequences and gene expression information have great potential to contribute to the understanding of the parasitism of Apicomplexa parasites, which could lead to better therapeutics and diagnostics, intensive international efforts have been made to conduct genome and transcriptome analyses of the parasites. The entire genome sequences of various malaria species, such as *Plasmodium falciparum* (Pf)*, P. vivax* (Pv) and *Toxoplasma gondii* (Tg), have been reported (1–3). Additionally, for the transcriptome analysis, cDNA libraries were constructed, and cDNA sequences were analyzed for Pf, *P. yoelii* (Py), Pv, *P. berghei* (Pb), Tg, *Cryptosporidium parvum* (Cp) and *Echinococcus multilocularis* (Em) (1–4). We constructed a series of full-length cDNA libraries using our oligo-capping method, which selectively captures mRNAs containing a cap structure (5). The obtained full-length cDNA information together with physical cDNA clones have been made publicly available from our database, Full-Parasites (http://fullmal.hgc.jp). Additionally, in the Comparasite sub-database, cDNA information is associated with putative mutually orthologous genes so that comparative genomic studies between different species are possible (6).

Recent massively parallel sequencing technologies, such as the Illumina GA sequencer system (7), have drastically reduced the sequencing cost per base. To complement common analytical methods, we developed several

*To whom correspondence should be addressed. Tel/Fax: +81 4 7136 3607; Email: ysuzuki@hgc.jp original procedures to utilize massively parallel sequence data in transcriptome analyses. We devised shotgun sequencing and genome-based assembly methods for determining the entire sequences of full-length cDNAs with a cost of less than one dollar per clone (8). We have also developed a method to generate numerous transcription start site (TSS) tags, which are short sequences immediately downstream of TSSs, by combining our oligo-capping method and Illumina GA technology [TSS-Seq; (9)].

In this update, in addition to expanding our full-length cDNA data set generated by the aforementioned methods, we have included two unique RNA-Seq (10) data sets (730 million mapped tags in total). One is the time-course expression profiling of bradyzoite differentiation for Tg. We used mixed RNA for the RNA sequencing, so that gene expression from the host and parasite were simultaneously analyzed (~95% of the RNA was human, and 5% was from parasites). The other RNA-Seq data set includes clinical samples from Pf. We extracted Pf RNA and the host human RNA from peripheral blood (~0.5–5% of RNA was from parasites). We expected that expression changes in Pf and immune responses in humans could be simultaneously monitored by analyzing the generated mixed RNA-Seq tags. In addition, by comparing the RNA sequences, we were able to analyze genetic variation among the Pf samples as well. We believe this type of data, which represents the dynamic nature of the transcriptome, will provide the most biologically relevant information now that many of the basic genetic elements have been identified and catalogued.

With the expanded cDNA contents, the enhanced functionality of the databases, and the new type of dynamic transcriptome data, we believe that the updated Full-Parasites is a useful data resource for understanding host–parasite interactions. We believe that integrative analyses of both causative parasites and host human cells will prove to be crucial for the eventual development of an effective method for preventing infectious diseases. Full-Parasites is accessible at http://fullmal.hgc.jp/.

EXPERIMENTAL PROCEDURES FOR THE PRODUCTION OF NEW DATA

RNA-Seq analysis of mixed RNA for Tg and Pf

For the RNA-Seq analysis of in-culture bradyzoite differentiation for Tg, the Tg ME49 strain was cultured in a monolayer of human foreskin fibroblasts (HFF) with Dulbecco's modified Eagle's medium (GIBCO). For the *in vitro* induction of bradyzoites, 1.5×10^6 HFF cells in each experimental condition were infected at an MOI of 0.5 followed by pre-culture for 24 h. Then culture media was replaced by RPMI 1640 medium (GIBCO) at pH 8.1 (adjusted with NaOH). The medium was exchanged every 2 days, and time course sampling was carried out over 144 h after the induction of differentiation. For each sample, total RNA was extracted from the infected cells using TRI reagent (Sigma). Approximately 20 µg of total RNA was extracted, and 1 µg was used to prepare the template for RNA-Seq using the RNA-Seq template preparation kit (Illumina), following the manufacturer's instructions. A single lane of 36-bp single-end sequencing (one-eighth split of a run) was performed, and at least 10 million sequence tags were generated for each sample. RNA-Seq tags were mapped to the reference genomes of humans (hg19) and parasites (ToxoDB Release 5.2), allowing two-base mismatches. No RNA tags were simultaneously mapped to the human and Tg genomes.

For the RNA-Seq analysis of clinical Pf samples, peripheral blood samples from patients infected by Pf in Indonesia, which were collected according to the protocol approved by the ethical committee of Sam Ratulangi University, were used. From 2.5 ml of the blood sample, total RNA was extracted using an RNA tube and an RNA extractor (PAX Gene). polyA+ RNA was selected and was used as a template for RNA-Seq. A single lane of 36-bp single-end sequencing was performed, and at least 10 million sequence tags were generated per sample. The generated RNA-Seq tags were mapped to the reference genomes of humans (hg19) and Pf (PlasmoDB Release 6.0), allowing two-base mismatches. As in the case of Tg, no RNA tags mapped to both human and Pf genomes.

New cDNA sequence data and TSS-Seq data

For the TSS-Seq analysis of Tg, the Tg ME49 strain was cultured and differentiated in the same way as for the RNA-Seq analysis. One hundred and forty-four hours after bradyzoite induction, bradyzoites were purified by Arabic gum density-gradient centrifugation, which is a method for separating cysts of Tg from the infected mammalian cells by multi-layer centrifugation in an Arabic gum solution. Briefly, 10 ml of Arabic gum solution having a specific gravity of 1.07 or 1.05 was layered in 50-ml tubes. Then, 16 ml of suspended infected cells, which was homogenized with a 23 G needle, was added and then centrifuged for 10 min at 800*g* at 20°C (11). Tachyzoites were purified by filtration through 5-mm pore membranes. Total RNA was extracted from the infected human cells using TRIzol reagent (Sigma). Approximately 200 µg of total RNA was extracted and was used as a template for TSS-Seq. Template preparation for TSS-Seq analysis was carried out as previously described (9). Briefly, the 5′- and 3′-adaptor sequences necessary for the Illumina GA sequencing were introduced as the 5′-end oligo during the RNA ligation and as the random hexamer primer during the first-strand cDNA synthesis, respectively. For each sample, a single lane of 36-bp single-end sequencing was performed. Five to 10 million TSS tags were generated and mapped to the respective reference genome sequences. The position to which the 5′-end of the Illumina GA sequence tag was mapped was defined as a putative TSS. Statistics for TSS tags are shown in Table 2. The mapped TSS tags were clustered to identify putative promoter regions. Details of the analysis of the identified TSS will be described elsewhere.

For the Sanger sequencing of cDNAs, oligo-cap cDNA libraries were constructed as previously described. Among the 10 000 randomly sequenced 5′-ESTs, non-redundant cDNAs were selected and subjected to shotgun sequencing using an Illumina GA system (8). On average, 800 cDNA clones were mixed, and 20 million shotgun tags were generated per pool. Genome-based assemblies were carried out as described in ref. (8).

DATABASE DESCRIPTIONS

Updated statistics and new functionalities of Full-Parasites

Since the last update in 2009, the data set of Full-Parasites has been extended to cover cDNA sequences for more Apicomplexa species, including *Babesia*, *Neospora*, *Eimeria* and *Theileria* species (Table 1). In the latest version, Full-Parasites contains 105 786 ESTs, of which 5925 cDNAs were selected for complete sequencing by shotgun sequencing coupled with genome-based assembly on an Illumina GAII system (8). TSS-Seq analysis was also carried out for different parts of the life cycles of Pf and Tg (Table 2).

Taking advantage of our unique full-length cDNA sequence data, multiple kinds of transcript-based annotations are possible in Full-Parasites. Various types of cDNA data are linked together to allow integrative interpretation of the data. To allow users to take advantage of this resource, Full-Parasites implements various types of viewers:

(i) Full-length cDNA Viewer: 5′-ESTs and the newly assembled complete sequences of the full-length cDNAs appear as a new track in addition to the original tracks in the genome viewer (left panel, Figure 1; http://fullmal.hgc.jp/). For each species, 300–2000 genes are covered. Because Apicomplexa parasites have approximately 4000–8000 genes (1–4), a significant proportion of the genes are represented in our database as full-length cDNAs. We primarily used genome-based assembly of the shotgun-sequenced cDNAs (800 cDNAs clones were mixed per pool) to expedite the complete sequencing. However, as an inevitable attribute of data obtained by the shotgun approach, some of the assembled sequences contain gaps or incompletely assembled regions (8). The assembly viewer was also constructed so that users can empirically understand and verify the quality and integrity of the assemblies used for the annotations (middle panel, Figure 1).

Table 1. Statistics for 5′-ESTs and complete cDNA sequences

Species	5′-ESTs	Completely sequenced cDNAs
Pf	9937	348
Pv	9633	2041
Py	11 581	311
Pb	2047	329
Cp	10 110	1066
Tg	7398	1830
Bb	12 286	n.d.
Be	7767	n.d.
Bc	10 769	n.d.
Nc	3456	n.d.
Et	7362	n.d.
Tp	13 440	n.d.

The genome sequences used for cDNA mapping are shown in the Database Glossary. Bb, *Babesia bovis*; Be, *Babesia equi*; Bc, *Babesia caballi*; Nc, *Neospora caninum*; Et, *Eimeria tenella*; Tp, *Theileria parva*. n.d., not determined.

Table 2. Statistics for TSS Seq tags

Species	Strain	Stage	Total tags	Mapped TSS tags	TSS positions
Tg	RH	Tachyzoite	6 801 945	2 591 387	85 750
Tg	ME49	Tachyzoite	12 101 228	2 484 257	242 889
Tg	ME49	Bradyzoite	8 418 271	357 792	67 091
Pf	3D7	Erythrocyte	4 870 527	673 313	239 284

(ii) Annotation Viewer: current annotations include homology (BLASTP), protein motifs (InterPro: http://www.ebi.ac.uk/interpro/ and Pfam: http://www.sanger.ac.uk/Software/Pfam/), GO term assignment (http://www.geneontology.org/index.shtml), hydropathy plotting (using the standard protocol), predictions of subcellular localization signals (PSORT: http://psort.hgc.jp/) and transmembrane domains (SOSUI: http://sosui.proteome.bio.tuat.ac.jp/sosuiframe0.html). For details of the functional annotation procedures, cut-offs and other parameters/criteria, see our website (http://fullmal.hgc.jp/comparas/Glossary.htm).

(iii) TSS Tag Viewer: the TSS viewer illustrates how TSS tags are distributed along the genomes of respective parasites. In the frame of the genome viewer, users can browse the locations of TSSs at different resolutions, from a genome-wide view to a single-base resolution. This function can be used to define the exact gene boundaries or to identify the exact positions of the upstream putative promoter regions. TSS information should also be useful for identifying previously overlooked transcripts, such as non-coding RNAs, which remain mostly undiscovered by standard gene predictions (12).

(iv) Phylogenic Analysis Viewer: with the expanded cDNA database, Comparasite, a sub-database of Full-Parasites for comparative studies of the Apicomplexa parasites, was also updated accordingly. Similar to the previous versions, users can search by inputting keywords (cDNA/gene ID), genomic positions, the presence or absence of various kinds of annotation features attached to annotated gene models or newly assembled complete full-length cDNA sequences. In the new version, users can search using evolutionary conservation patterns. For example, users can search by specifying the base-substitution rate, the rate of synonymous/non-synonymous substitutions, and the shape of the phylogenetic trees (right panel, Figure 1; http://fullmal.hgc.jp/cgi-bin/evolution.cgi).

In general, although the contents of the cDNA data were expanded and several new features and new viewers

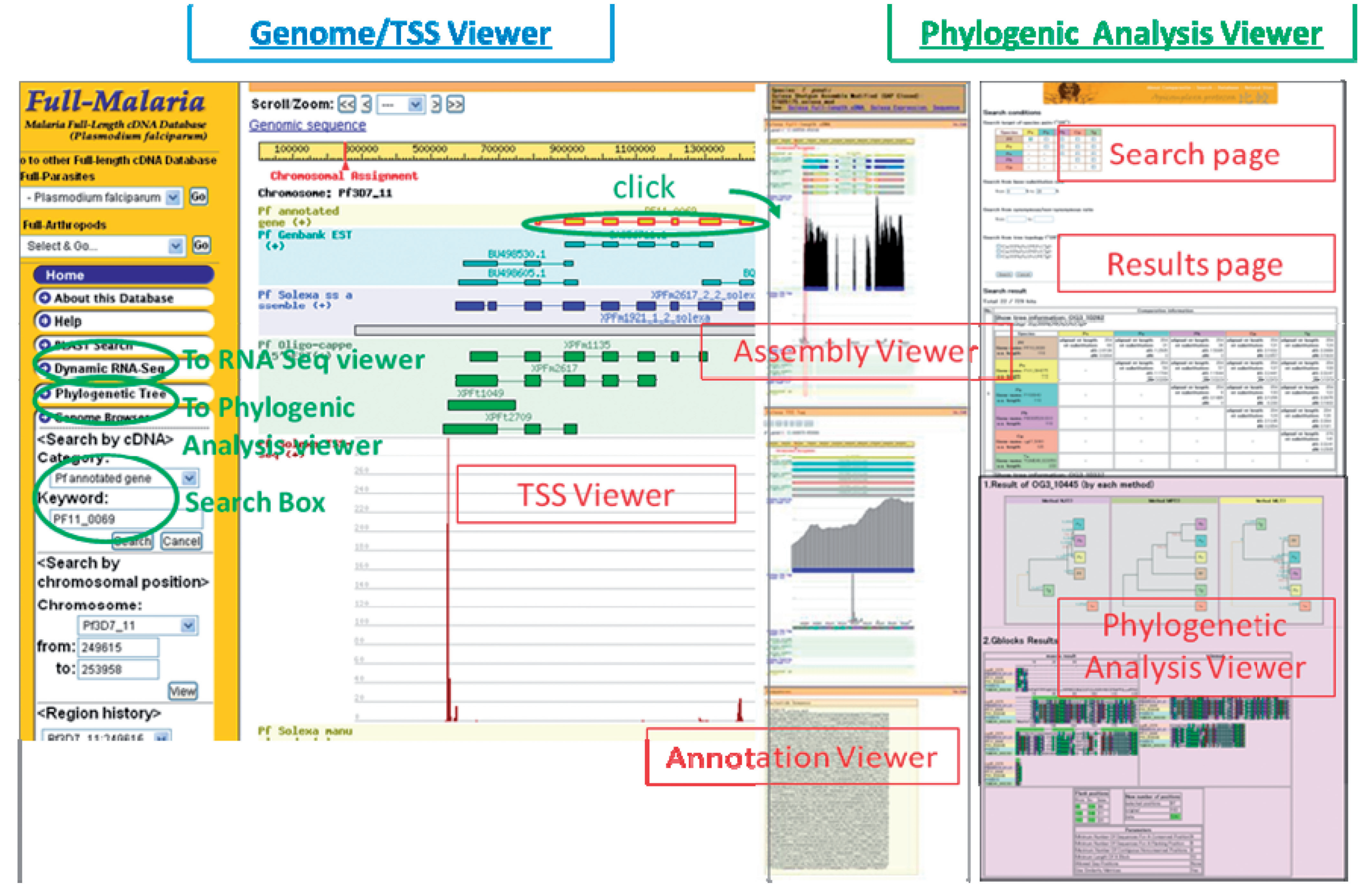


Figure 1. Updated Browser of Full-Parasites. Screen shots of the Genome Browser (left panel) implemented with the TSS Viewer (middle panel), the Annotation Viewer (middle panel) and the Phylogenetic Analysis viewer (right panel). To search the database, specify the species and gene name/cDNA ID in the boxes indicated by the green circle at the top of the page (http://fullmal.hgc.jp/). To search for genes having particular evolutionary conservation patterns, specify the shape of the phylogenetic trees or base substitution rate at the top page of the Phylogenetic Analysis Viewer (http://fullmal.hgc.jp/cgi-bin/evolution.cgi). To search for genes having particular expression patterns, follow the link to the Dynamic RNA-Seq Viewer (http://fullmal.hgc.jp/cgi-bin/dynamic.cgi). These pages are also linked from the Annotation Viewer. Details of the search conditions, legends for coloring and items are described in the Database Glossary (http://fullmal.hgc.jp/docs/glossary.html).

were introduced, the overall look and feel of the viewer remains unchanged from the previous version to avoid confusion.

Mixed RNA sequencing data for analyzing human–parasite transcriptome interactions

To analyze the interactions among transcripts of infecting parasites and infected human cells, we applied an RNA-Seq analysis of the mixed RNAs isolated from infected human cells. The current version comprised two data sets containing a total of 730 million mapped RNA-Seq tags (Table 3). The first data set is taken from Tg which were differentiated into bradyzoites in culture. We extracted human and parasite RNAs during the time course of bradyzoite differentiation. The collected RNAs were subjected to RNA-Seq analysis. At least 10 million RNA-Seq tags were generated for each sample. Of these RNA-Seq tags, ~95% originated from human RNA, and the rest were from Tg. Various gene expression patterns were observed during the differentiation process in both humans and Tg (details of the analysis will be published elsewhere).

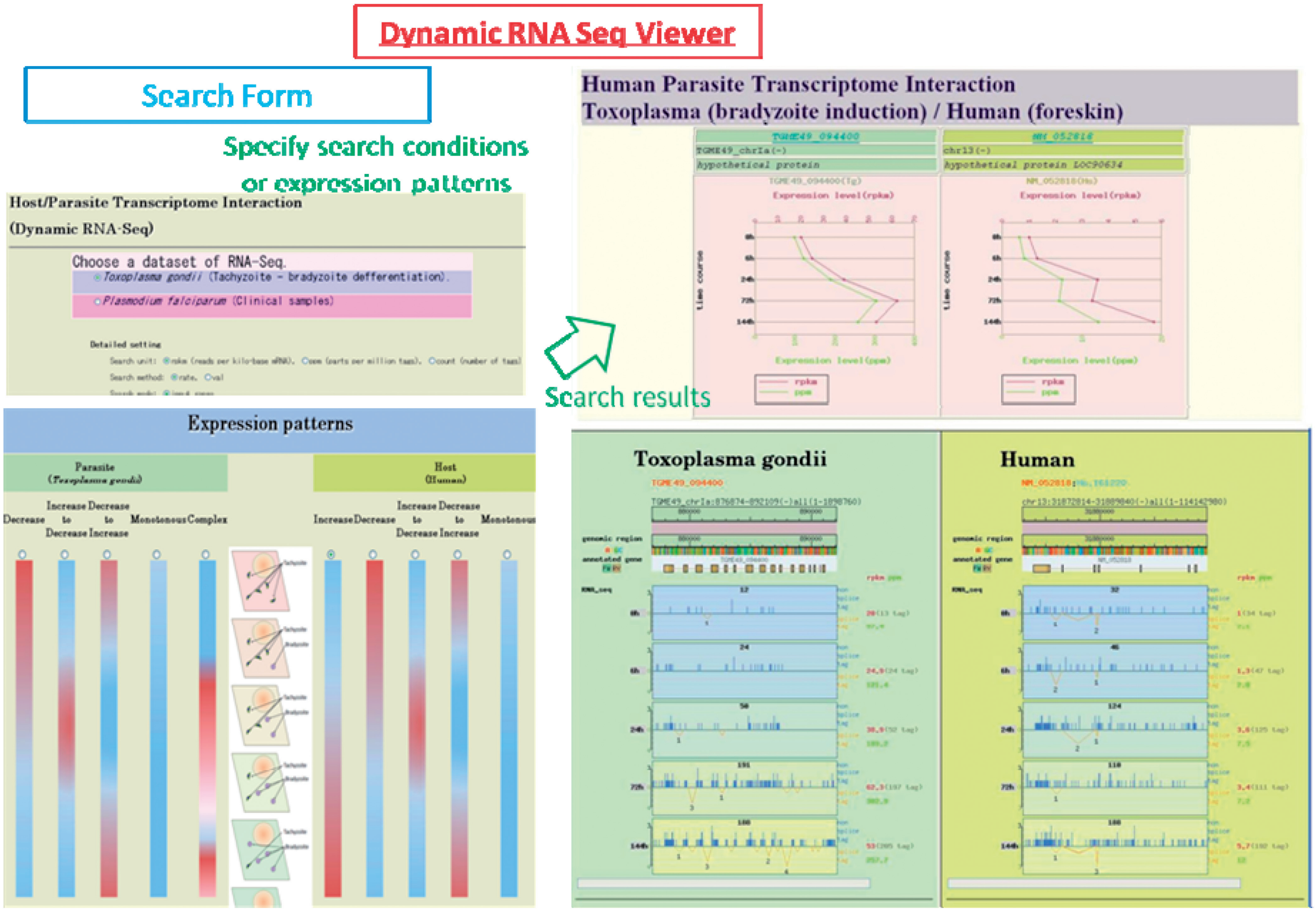
In the Dynamic RNA-Seq Viewer (http://fullmal.hgc.jp/cgi-bin/dynamic.cgi), users can search for human and Tg genes by specifying the fold change in the gene expression level relative to that at time zero. Users can also use the absolute expression levels, which were evaluated by tag counts for the search. A search using overall gene expression patterns is also possible. As exemplified in the following section, users can search by whether the gene expression monotonically increased or decreased or whether there was an inflection point (details of how to set the search conditions, e.g. how to specify the standard point to define the relative expression levels, are described on the help page of the Dynamic RNA-Seq Viewer). Search results are linked to the cDNA viewer so that users can directly obtain the functional information for the genes that show particular expression patterns (left panel, Figure 2).

The second RNA-Seq data set is from Pf field samples. For this data set, we used blood samples collected from 31 patients infected by Pf in Indonesia. At least 10 million RNA-Seq tags for each sample were generated. Among the collected RNA-Seq tags, ~0.5–5% was from Pf,

Table 3. Statistics for dynamic RNA-Seq tags

Species	Data sets	Total mapped tags	Average frequency of parasite tags, %	Mapped tags	Average represented gene (>0 ppm)	Average represented gene (>1 ppm)	Average cSNPs detected[a]
Tg	16	234 121 334	2.1	4 076 308	4207	3409	n.d
Human				230 045 026	19 426	16 151	n.d.
Pf	31	501 067 025	2.3	11 071 543	2952	2917	161
Human				489 995 482	18 797	14 464	n.d.

[a] cSNPs were detected using Genome Studio (Illumina) with default settings.
n.d., not determined.



Host/Parasite Transcriptome Interaction (Dynamic RNA-Seq)

Choose a dataset of RNA-Seq.
● *Toxoplasma gondii* (Tachyzoite - bradyzoite defferentiation).
○ *Plasmodium falciparum* (Clinical samples)

Detailed setting
Search unit: ● rpkm (reads per kilo-base mRNA), ○ ppm (parts per million tags), ○ count (number of tags)
Search method: ● rate, ○ val

Figure 2. Search Page and Results Page of the Dynamic RNA-Seq Viewer. The search page (left panel) and search results page (right panel) of the Dynamic RNA-Seq Viewer. To search the database, specify the search conditions or expression patterns at the top of the page of the Dynamic RNA-Seq viewer (http://fullmal.hgc.jp/cgi-bin/dynamic.cgi); also specify whether the expression patterns should be considered to be values relative to the indicated time point/patient or to be absolute tag counts and whether the expression patterns should show typical patterns or specified patterns.

depending on the parasitemia (infection rate of parasites in erythrocytes). The rest were from humans, which gives information on the gene expression changes in human peripheral blood cells induced by malaria infection. Again, various expression patterns were observed from different host–parasite pairs. Users can search for genes showing particular expression patterns in different patients and infection stages. The basic search options are similar to those for the Tg data set. Clinical information for malaria symptoms, such as body temperature and suspected date of malaria infection, are also presented. In addition to the dynamic nature of human–parasite transcriptomes, we were also able to identify a large number of genetic variations, which were detected as cSNPs (cDNA SNP), in the field Pf samples compared to the reference genome sequence (Table 3). Users can also search for these genetic variations using our database. Details of the biological analysis of the tag information will be published elsewhere. We expect that further extensive transcriptome analysis using a larger number of clinical samples will provide useful information for understanding the clinical symptoms of malaria infections in Indonesia.

Search examples

For an example of a search, follow these steps (Figure 1): Full-Parasites top; select the species, *P. falciparum,* and specify the ‘Annotated gene ID’ as ‘PF11_0069′ (in ‘Search Box’ shown in Figure 1). Evolutional conservation patterns (Phylogenic Analysis Viewer) and expression patterns of the gene (Dynamic RNA-Seq viewer) can also be followed from the Annotation Viewer, which are linked from the model transcript (indicated by a green circle, Figure 1).

Glossary, data and clone repository

A detailed user manual and a list of technical terms, definitions and parameters for the annotations are described in the ‘Glossary and Experimental Procedure’ sections of our websites (http://fullmal.hgc.jp/docs/glossary.html; http://fullmal.hgc.jp/docs/procedure.html). Users can follow the links for further information on each item displayed there. Statistics for the current database are also presented in the statistics section (http://fullmal.hgc.jp/docs/statistics.html). All of the short read sequences used for the database have been deposited in the NCBI Short Read Archives (http://www.ncbi.nlm.nih.gov/Traces/sra/sra.cgi). Newly generated RNA-Seq data have also been registered under the following accession numbers: DRA000224–DRA000273. The raw sequence data are publicly and freely available from the download site in our database (left frame in the middle in the top page). Additionally, the cDNA clones registered in the database are freely available.

CONCLUSIONS AND FUTURE PERSPECTIVES

Herein, we described an update of our Full-Parasites database with extensive cDNA data and a new type of dynamic RNA-Seq data. To visualize the newly generated short read sequences, we implemented a new version of the genome viewer and the RNA-Seq viewer. Particularly for the new type of RNA-Seq data sets, further enrichment of the data from additional clinical samples and more laboratory strain data during different life cycle stages and in different culture conditions is also being explored. Because it focuses on the dynamic nature of the transcriptome data and is based on various types of cDNA analyses, our database is different from other parasite databases, such as PlasmoDB (http://www.plasmodb.org/), CryptoDB (http://cryptodb.org/) and ToxoDB (http://www.toxodb.org/), whose main focus is on the static annotation of gene components. Through complementary use of our database and others, we believe that we will be able to lay a strong foundation for understanding how Apicomplexa parasites interact with host transcriptomes and achieve such complex life cycles with a limited number of genes.

ACCESSION NUMBER

DRA000224–DRA000273

ACKNOWLEDGEMENTS

This work has been long conducted by Dr Junichi Watanabe, who recently passed away. We are grateful to Ms Etsuko Sekimori and Keiko Toya for excellent programming work. The full-length cDNA libraries were produced by Ms Kazumi Abe. Large-scale sequencing was performed by Ms Kiyomi Imamura, Rie Nishizawa and Mari Tsubaki. We are thankful to Mr Fah Sathirapongsasuti for critical reading of the manuscript. This database was constructed and maintained by a Grant-in-Aid for Publication of Scientific Research Results from the Japan Society for Promotion of Science and JST.

FUNDING

RNA sequencing of malaria patients and *Toxoplasma gondii*-infected cells was funded by the grant Asia-Africa S & T Strategic Cooperation Promotion Program by the Special Coordination Funds for Promoting Science & Technology from the Ministry of Education, Culture, Sports, Science and Technology of Japan. Funding for open access charge: Asia-Africa S & T Strategic Cooperation Promotion Program by the Special Coordination Funds for Promoting Science and Technology from the Ministry of Education, Culture, Sports, Science and Technology of Japan.

Conflict of interest statement. None declared.

REFERENCES

1. Carlton,J.M., Angiuoli,S.V., Suh,B.B., Kooij,T.W., Pertea,M., Silva,J.C., Ermolaeva,M.D., Allen,J.E., Selengut,J.D., Koo,H.L. *et al.* (2002) Genome sequence and comparative analysis of the model rodent malaria parasite Plasmodium yoelii yoelii. *Nature*, **419**, 512–519.
2. Gardner,M.J., Hall,N., Fung,E., White,O., Berriman,M., Hyman,R.W., Carlton,J.M., Pain,A., Nelson,K.E., Bowman,S. *et al.* (2002) Genome sequence of the human malaria parasite Plasmodium falciparum. *Nature*, **419**, 498–511.
3. Gajria,B., Bahl,A., Brestelli,J., Dommer,J., Fischer,S., Gao,X., Heiges,M., Iodice,J., Kissinger,J.C., Mackey,A.J. *et al.* (2008) ToxoDB: an integrated Toxoplasma gondii database resource. *Nucleic Acids Res.*, **36**, D553–D556.
4. Aurrecoechea,C., Brestelli,J., Brunk,B.P., Dommer,J., Fischer,S., Gajria,B., Gao,X., Gingle,A., Grant,G., Harb,O.S. *et al.* (2009) PlasmoDB: a functional genomic database for malaria parasites. *Nucleic Acids Res.*, **37**, D539–D543.
5. Suzuki,Y. and Sugano,S. (2003) Construction of a full-length enriched and a 5’-end enriched cDNA library using the oligo-capping method. *Methods Mol. Biol.*, **221**, 73–91.
6. Wakaguri,H., Suzuki,Y., Katayama,T., Kawashima,S., Kibukawa,E., Hiranuka,K., Sasaki,M., Sugano,S. and Watanabe,J. (2009) Full-Malaria/Parasites and Full-Arthropods: databases of full-length cDNAs of parasites and arthropods, update 2009. *Nucleic Acids Res.*, **37**, D520–D525.
7. Bentley,D.R., Balasubramanian,S., Swerdlow,H.P., Smith,G.P., Milton,J., Brown,C.G., Hall,K.P., Evers,D.J., Barnes,C.L., Bignell,H.R. *et al.* (2008) Accurate whole human genome

sequencing using reversible terminator chemistry. *Nature*, **456**, 53–59.

8. Kuroshu,R.M., Watanabe,J., Sugano,S., Morishita,S., Suzuki,Y. and Kasahara,M. (2010) Cost-effective sequencing of full-length cDNA clones powered by a de novo-reference hybrid assembly. *PLoS One*, **5**, e10517.
9. Tsuchihara,K., Suzuki,Y., Wakaguri,H., Irie,T., Tanimoto,K., Hashimoto,S., Matsushima,K., Mizushima-Sugano,J., Yamashita,R., Nakai,K. *et al.* (2009) Massive transcriptional start site analysis of human genes in hypoxia cells. *Nucleic Acids Res.*, **37**, 2249–2263.
10. Wang,Z., Gerstein,M. and Snyder,M. (2009) RNA-Seq: a revolutionary tool for transcriptomics. *Nat. Rev. Genet.*, **10**, 57–63.
11. Nakabayashi,T. and Motomura,I. (1968) A method for separating cysts of Toxoplasma gondii from the infected mouse brains by multi-layer centrifugation with gumarabic solution. *Tropical Medicine*, **10**, 72–80.
12. Wakaguri,H., Suzuki,Y., Sasaki,M., Sugano,S. and Watanabe,J. (2009) Inconsistencies of genome annotations in apicomplexan parasites revealed by 5'-end-one-pass and full-length sequences of oligo-capped cDNAs. *BMC Genomics*, **10**, 312.