



## Data in Brief

# Technical data of the transcriptomic analysis performed on tsetse fly symbionts, *Sodalis glossinidius* and *Wigglesworthia glossinidia*, harbored, respectively by non-infected, *Trypanosoma brucei gambiense* infected and self-cured *Glossina palpalis gambiense* tsetse flies

Anne Geiger<sup>a,\*</sup>, Bernadette Tchicaya<sup>a</sup>, Pascal Rihet<sup>b,c</sup><sup>a</sup> UMR 177, IRD-CIRAD, CIRAD TA A-17/G, Campus International de Baillarguet, 34398 Montpellier Cedex 5, France<sup>b</sup> UMR1090 TAGC, INSERM, Marseille F-13288, France<sup>c</sup> Aix-Marseille University, Marseille F-13288, France

## ARTICLE INFO

## Article history:

Received 1 April 2015

Accepted 6 April 2015

Available online 14 April 2015

## Keywords:

Genes differentially expressed

Microarrays

*Glossina palpalis gambiense**Trypanosoma brucei gambiense*

Tsetse fly infection

Sleeping sickness

## ABSTRACT

Microarray is a powerful and cheap method to identify and quantify gene expression in particular in a mix of total RNA extracted from biological samples such as the tsetse fly gut, including several organisms (here, the fly tissue and the intestinal microorganisms). Besides, biostatistics and bioinformatics allow comparing the transcriptomes from samples collected from differently treated flies, and thus to identify and quantify differential expressed genes. Here, we describe in details a whole microarray transcriptome dataset produced from tsetse flies symbionts, *Sodalis glossinidius* and *Wigglesworthia glossinidia*. The tsetse fly midguts were sampled at key steps of tsetse fly infection by trypanosomes, 3-day and 10-day sampling times to target differentially expressed genes involved, respectively, in early events associated with trypanosome entry into the midgut and with the establishment of infection; 20 days to target the genes involved in events occurring later in the infection process. We describe in detail the methodology applied for analyzing the microarray data including differential expression as well as functional annotation of the identified symbiont genes. Both the microarray data and design are available at <http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE48360>; <http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE48361>; <http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE55931>.

© 2015 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

## Specifications

Organism/cell line/tissue	Symbionts, <i>Sodalis glossinidius</i> and <i>Wigglesworthia glossinidia</i> , from <i>Glossina palpalis gambiense</i> midgut
Sex	Female tsetse flies
Sequencer or array type	Custom-made 60-mers oligonucleotide microarrays
Data format	Raw and analyzed (microarray)
Experimental factors	Symbiont transcriptomes from tsetse flies infected by trypanosomes, self-cured and from control tsetse flies
Experimental features	<i>G. p. gambiense</i> midguts were sampled at three times (3, 10 and 20 days) post-trypanosome-infected and non-infected bloodmeal uptake by the tsetse flies. The transcriptomes of the tsetse fly symbionts, <i>S. glossinidius</i> , and <i>W. glossinidia</i> , issued from these guts, were compared as to identify genes that are differentially expressed.

(continued)

## Specifications

Consent	Experiments on animals reported in this article were conducted according to internationally recognized guidelines and were approved by the Ethics Committee on Animal Experiments and the Veterinary Department of the Centre International de Recherche Agronomique pour le Développement (CIRAD), Montpellier-France.
Sample source location	Insectary from CIRAD, Montpellier-France

## Direct link to deposited data

<http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE48360>
<http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE48361>
<http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE55931>

\* Corresponding author at: UMR 177, IRD-CIRAD, CIRAD TA A-17/G, Campus International de Baillarguet, 34398 Montpellier Cedex 5, France. Tel.: +33 4 67 59 39 25; fax: +33 4 67 59 38 94.

E-mail address: [anne.geiger@ird.fr](mailto:anne.geiger@ird.fr) (A. Geiger).

## Experimental design, materials and methods

### Experimental design

Tsetse flies of the sub-species *Glossina palpalis gambiensis* were infected by trypanosomes of the sub-species *Trypanosoma brucei gambiense*. At key steps of flies' infection, 3, 10, and 20 days, midguts of flies were dissected and total RNA was extracted in order to further analyze the transcriptome of tsetse fly symbionts, *Sodalis glossinidius* and *Wigglesworthia glossinidia*. The sampling days are chosen to target events associated respectively, i) with trypanosome entry into the midgut, ii) with the establishment of infection, and iii) with the late stages of the infection process. Fig. 1 shows the general experimental design.

### Materials and methods

#### Experimental infection of *G. p. gambiensis* by *T. b. gambiense*

Insectary *G. p. gambiensis* flies from CIRAD, Montpellier, were *T. b. gambiense* infected experimentally according to the protocol reported by Geiger et al. [1] and Hamidou Soumana et al. [2–4]. Strain of *T. b. gambiense* S7/2/2 (isolated in 2002 from a HAT patient diagnosed in the sleeping sickness focus of Bonon, Ivory Coast [5]) was injected intraperitoneally into balb/cj mice. After the parasitemia has reached  $15\text{--}25 \times 10^7$  parasites/ml, teneral flies were fed on these infected mice. This group of flies was then separated into three sub-groups a, b, and c. Three days after feeding, four biological replicates, each of the seven flies, were randomly selected from the sub-group a; they were noticed (S3 for “stimulated-sampled at day3”). Ten days after feeding, the flies of the sub-group b were tested for the presence/absence of trypanosomes in their anal drop and separated into two “sub-sub-groups”,

one noticed “I10” (flies fed on infected mice and that were shown to be infected, sampled at Day 10 post-feeding), the second noticed “NI10” [flies fed on infected mice and that were shown to be non-infected (= refractory flies), sampled at Day 10 post-feeding]. Twenty days after feeding the sub-group c was processed as was the sub-group b; the corresponding “sub-sub-groups” were noticed “I20” and “NI20”. From each “S3”, “I20”, and “NI20”, 4 biological replicates were constituted each of 7 flies randomly sampled. For I10 and NI10: 4 replicates were constituted of 3 flies because of the low infection prevalence. Finally, a group of flies was fed on non-infected mice, of which four replicates, each of 7 flies, were constituted, three days after feeding, and noticed “NS3” (for non-stimulated = control flies).

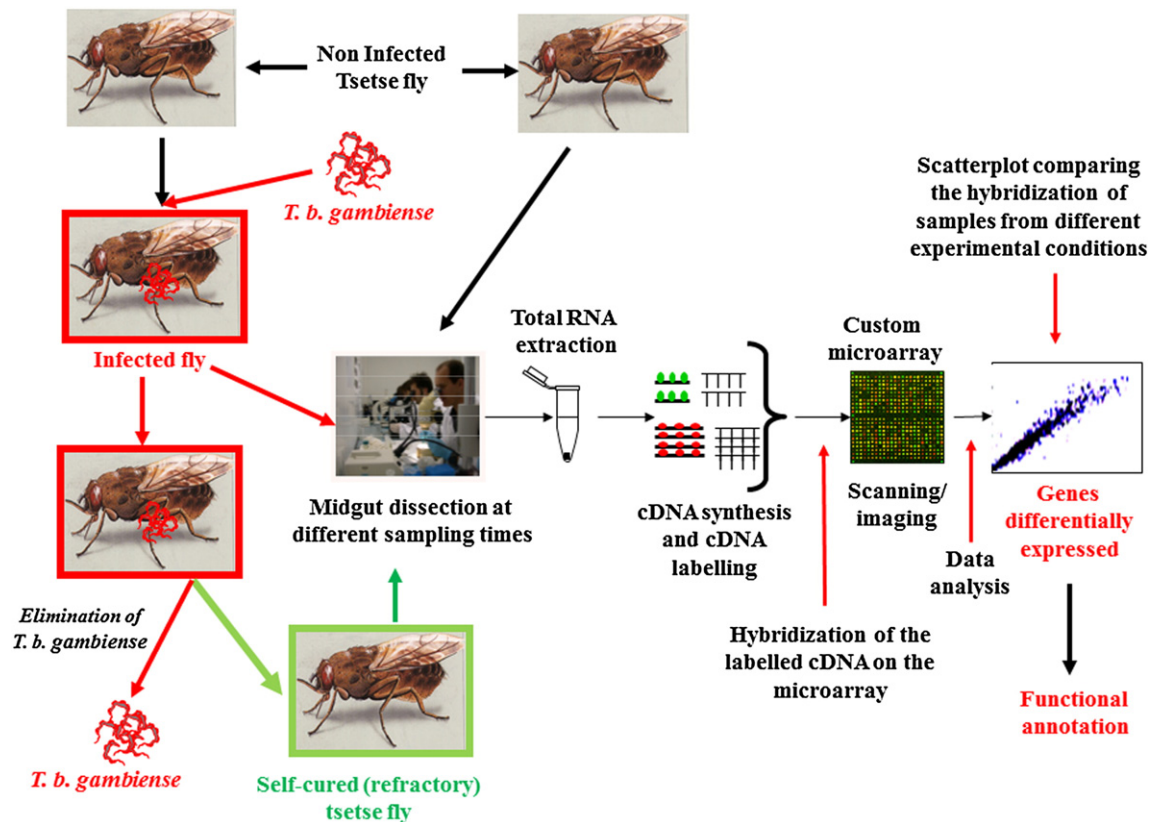
#### Fly infection monitoring process

As mentioned, flies fed on infected mice and sampled at Day 10 and Day 20 were controlled for the presence or absence of trypanosomes in their anal drops. This was performed on chelex-extracted DNA [6] from the anal drops and the presence of trypanosomes was assessed by PCR using TBR1 and TBR2 primers [7]. When anal drops were PCR positive for the presence of trypanosomes, it indicates midgut infections. When PCR tests were negative, flies had self-cured the infection.

#### RNA extraction

Flies from the different biological repeats (from “S3”, “NS3”, ...) were then dissected separately and the midguts were collected in RNA lysis buffer (Ambion) for further RNA extraction.

RNA was extracted from the midguts of each biological replicate using TRIzol reagent (Gibco-BRL, France). High quality of RNA sample was checked on an Agilent RNA 6000 Bioanalyzer and the RNA quantification was performed using the corresponding Nano kit (Agilent Technologies, France).



**Fig. 1.** General experiment design. Midgut of *G. p. gambiensis* was sampled at three times post-*T. b. gambiense* infected bloodmeal: 3, 10, and 20 days. For each time points, 4 biological replicates of seven or three (for the I10 and NI10 samples only) midguts were constituted and further analyzed for *Sodalis* or *Wigglesworthia* transcriptome. Total RNA was produced from each biological replicate, and reverse transcribed into cDNA that was then labeled and hybridized onto *Sodalis* or *Wigglesworthia* custom-made microarrays. Genes differentially expressed between the different conditions were further analyzed and annotated.

### Custom-made 60-mers oligonucleotide microarrays

The tsetse fly symbiont custom-made density arrays (8 × 15 K format) were designed with 60-mer oligos specific to:

\* For *Sodalis* [2,3]: genes of the *S. glossinidius* chromosome (NCBI RefSeq: NC\_007712.1; GenBank accession number AP008232), and genes of the *Sodalis* four plasmids pSG1 (NCBI RefSeq: NC\_007183.1), pSG2 (NCBI RefSeq: NC\_007184.1), pSG3 (NCBI RefSeq: NC\_007186.1), and pSG4 (NCBI RefSeq: NC\_007187.1) [8,9]. Four unique probes were designed for each gene.

\* For *Wigglesworthia* [4]: genes of the *W. glossinidia* chromosome (from *Glossina morsitans morsitans*) (NCBI Reference Sequence: NC\_016893.1) [10]. Ten different probes were used for each gene.

To avoid cross-hybridization with non-target genes, for *Sodalis* and *Wigglesworthia* custom-microarrays, probes were selected only when they correspond to unique sequences.

The details of the *Sodalis* and *Wigglesworthia* array design, sample description, and expression data are available at Gene Expression Omnibus (GEO) under accession numbers respectively, GPL17347 and GSE48361 for *Sodalis*:

(<http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE48360>;  
<http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE48361>) [2,3],  
 and GPL18427 and GSE55931 for *Wigglesworthia*:  
 (<http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE55931>) [4].

### Preparation of cDNA and hybridization on *Sodalis* and *Wigglesworthia* custom-microarray

Microarray experiments were performed at the TAGC core facility (<http://tagc.univ-mrs.fr/>) for *Sodalis*, and at Hybrigenics platform (Clermont-Ferrand, France) for *Wigglesworthia*.

*Sodalis* cDNA labeling with Cy3 dCTP was done with 5 µg of total RNA using the ChipShot direct labeling and clean-up system kit (Promega). Samples were then hybridized onto the *Sodalis* custom-microarrays made from *S. glossinidius* Genome. Labeling of *Wigglesworthia* cDNA was performed with Cy3 dCTP and 100 ng of total RNA using the Low Input Quick Amp Labeling Kit One-Color (Agilent Technologies, France). cDNA samples were then hybridized onto the custom-microarrays made from *W. glossinidia* genome.

Hybridization was performed, for both custom-microarrays types, at 65 °C for 17 h at 60 rpm.

### Microarray data analyses

Lowest normalization was used for within-array normalization. Quantile normalization was used to make the density distributions similar across arrays [11]. Only one expression value was then assigned to each biological replicate by averaging the normalized expression values through Cy3 signal intensities. The pictures of microarray data scanned with an Agilent microarray scanner (Agilent Technologies) were extracted with the software (version 10.5.1.1) Agilent Feature Extraction. When in at least three of the four replicates, data show with expression levels greater than the background noise, then they were selected for further analyses. In the case of *Sodalis* transcriptome analyses, statistics was performed using the TIGR MeV (MultiExperiment Viewer) v4.5 software (<http://www.tm4.org/mev.html>). A two-way ANOVA was used to analyze, simultaneously, the effect of infection and of the time course post-feeding on trypanosome-infected mice on gene expression [12]. *p*-Values were calculated after 10,000 permutations, and multiple testing was controlled [13,14] using a FDR of 5%.

Unsupervised hierarchical clustering was applied to median-centered data, using the Cluster and TreeView programs (average linkage clustering using Pearson's correlation as the metric distance) to investigate relationships between samples and between genes.

One-way analysis of variance was applied to identify *Sodalis* genes differentially expressed between infection self-cured and control flies. A FDR of 5% was used for differential expression threshold [14].

Regarding *Wigglesworthia*, background adjustment, quantile normalization of data [11,15], log-transformation, and gene clustering analyses, were performed with GeneSpring GX (version 12.0, Agilent Technologies). A *t*-test was used for statistics [16]. A *p*-value below 0.05 indicates significant differences between groups (*Wigglesworthia* from 3 day stimulated flies versus 3 day control flies, *Wigglesworthia* from 10 day infected flies versus 10 day self-cured flies, and finally, *Wigglesworthia* from 20 day infected flies versus 20 day self-cured flies).

### Functional annotation of differentially expressed genes

Regarding *Sodalis* transcriptomes, functional annotation of differentially expressed genes (DEGs) was performed using DAVID software [17]. It was used to assess whether specific biological functions or pathways were overrepresented among the DEGs, based on gene ontology (GO) terms and on the Kyoto Encyclopaedia of Genes and Genomes (KEGG) pathways (<http://www.genome.jp/kegg/>). A score based on Fisher's exact test reflected the probability that the prevalence of a particular term within a cluster was a simple matter of chance or not. The *p*-values were corrected to account for multiple testing [13]; a *p*-values lower than 0.05 was considered significant.

As concerns the *Wigglesworthia* samples, GO was performed using the GeneSpring database, and *Wigglesworthia* gene expression data were also analyzed by principle component analysis (PCA) [18,19] performed with GeneSpring on infected (or stimulated) vs. non-infected (or non-stimulated) conditions at the different experimental infection times.

### Quantitative real-time PCR (qPCR) analysis

cDNA was synthesized from 5 µg of total RNA from each biological replicates using random hexamers and Superscript II reverse-transcriptase (Invitrogen, France). qPCR was then tested on some *Sodalis* genes that were differentially expressed in microarray experiments between the different groups of flies analyzed to confirm the results. Primers specific to the chosen genes were designed using Primer-Blast software (<http://www.ncbi.nlm.nih.gov/tools/primer-blast/>). qPCR reactions were then performed in an Mx3005P QPCR System (Agilent Technologies) using the Brilliant II Sybrgreen qPCR Kit (Agilent technologies) with 2 µl of cDNA in a 25-µl total volume. qPCR was analyzed using the qPCR Stratagene MxPro 3005P data analysis software. Efficiencies of the PCR reactions for each primer pair were calculated using ten-fold dilutions of fly gut-extracted cDNA [20]. Melting curve analysis was performed to check the specificity of the PCR reaction and to verify the amplification efficiency. Relative quantification was calculated with the  $2^{-\Delta\Delta C(\tau)}$  method as described by Livak and Schmittgen [21].

### Conflict of interest statement

There is no conflict of interest with respect to funding or any other issue.

### Acknowledgments

The authors thank the “Région Languedoc-Roussillon – Appel d'Offre Chercheur d'Avenir 2011”, the “Service de Coopération et d'Action Culturelle de l'Ambassade de France au Niger”, and the “Institut de Recherche pour le Développement” for their financial support.

### References

- [1] A. Geiger, et al., Vector competence of *Glossina palpalis gambiensis* for *Trypanosoma brucei* s.l. and genetic diversity of the symbiont *Sodalis glossinidius*. *Mol. Biol. Evol.* 2014) 102–109.
- [2] I. Hamidou Soumana, et al., The transcriptional signatures of *Sodalis glossinidius* in the *Glossina palpalis gambiensis* flies negative for *Trypanosoma brucei gambiense*

- contrast with those of this symbiont in tsetse flies positive for the parasite: possible involvement of a *Sodalis*-hosted prophage in fly *Trypanosoma* refractoriness? *Infect. Genet. Evol.* 24 (2014) 41–56, <http://dx.doi.org/10.1016/j.meegid.2014.03.005>.
- [3] I. Hamidou Soumana, et al., Identification of overexpressed genes in *Sodalis glossinidius* inhabiting trypanosome-infected self-cured tsetse flies. *Front. Microbiol.* 5 (2014) 255, <http://dx.doi.org/10.3389/fmicb.2014.00255>.
- [4] I. Hamidou Soumana, et al., Comparative gene expression of *Wigglesworthia* inhabiting non-infected and *Trypanosoma brucei gambiense*-infected *Glossina palpalis gambiense* flies. *Front. Microbiol.* 5 (2014) 620, <http://dx.doi.org/10.3389/fmicb.2014.00620>.
- [5] S. Ravel, et al., Cyclical transmission of *Trypanosoma brucei gambiense* in *Glossina palpalis gambiense* displays great differences among field isolates. *Acta Trop.* 100 (2006) 151–155.
- [6] S. Ravel, et al., Monitoring the developmental status of *Trypanosoma brucei gambiense* in the tsetse fly by means of PCR analysis of anal and saliva drops. *Acta Trop.* 88 (2003) 161–165.
- [7] D.R. Moser, et al., Detection of *Trypanosoma congolense* and *Trypanosoma brucei* subspecies by DNA amplification using the polymerase chain reaction. *Parasitology* 99 (1989) 57–66.
- [8] A.C. Darby, et al., Extrachromosomal DNA of the symbiont *Sodalis glossinidius*. *J. Bacteriol.* 187 (2005) 5003–5007.
- [9] H. Toh, et al., Massive genome erosion and functional adaptations provide insights into the symbiotic lifestyle of *Sodalis glossinidius* in the tsetse host. *Genome Res.* 16 (2006) 149–156.
- [10] L. Akman, et al., Genome sequence of the endocellular obligate symbiont of tsetse, *Wigglesworthia glossinidia*. *Nat. Genet.* 32 (2002) 402–407, <http://dx.doi.org/10.1038/ng986>.
- [11] B.M. Bolstad, et al., A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. *Bioinformatics* 19 (2003) 185–193.
- [12] M.K. Kerr, et al., Analysis of variance for gene expression microarray data. *J. Comput. Biol.* 7 (2000) 819–837.
- [13] Y. Benjamini, et al., Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. Ser. B* 57 (1995) 289–300.
- [14] A. Reiner, et al., Identifying differentially expressed genes using false discovery rate controlling procedures. *Bioinformatics* 19 (2003) 368–375.
- [15] G.K. Smyth, et al., Use of within-array replicate spots for assessing differential expression in microarray experiments. *Bioinformatics* 21 (2005) 2067–2075.
- [16] X.Y. Qin, et al., Identification of novel low-dose bisphenol A targets in human foreskin fibroblast cells derived from hypospadias patients. *PLoS One* 7 (2012) e36711, <http://dx.doi.org/10.1371/journal.pone.0036711>.
- [17] G. Dennis Jr., et al., DAVID: database for annotation, visualization, and integrated discovery. *Genome Biol.* 4 (2003) P3.
- [18] J.M. Claverie, Computational methods for the identification of differential and coordinated gene expression. *Hum. Mol. Genet.* 8 (1999) 1821–1832, <http://dx.doi.org/10.1093/hmg/8.10.1821>.
- [19] A. Butte, The use and analysis of microarray data. *Nat. Rev. Drug Discov.* 1 (2002) 951–960, <http://dx.doi.org/10.1038/nrd961>.
- [20] I. Hamidou Soumana, et al., Population dynamics of *Glossina palpalis gambiense* symbionts, *Sodalis glossinidius*, and *Wigglesworthia glossinidia*, throughout host-fly development. *Infect. Genet. Evol.* 13 (2013) 41–48, <http://dx.doi.org/10.1016/j.meegid.2012.10.003>.
- [21] K.J. Livak, T.D. Schmittgen, Analysis of relative gene expression data using real-time quantitative PCR and the  $2^{-\Delta\Delta CT}$  method. *Methods* 25 (2001) 402–408.