Supplementary material

# Dual RNAseq highlights the kinetics of skin microbiome and fish host responsiveness to bacterial infection

J. Le Luyer1,\*, Q. Schull1, 2, P. Auffret1, P. Lopez1,3, M. Crusot1,4, C. Belliard1, C. Basset1, Q. Carradec5, J. Poulain5 , S. Planes6, D. Saulnier1

1. Ifremer, IRD, Institut Louis‐Malardé, Univ Polynésie française, EIO, F‐98719 Taravao, Tahiti, Polynésie Française

2. MARBEC, Univ. Montpellier, Ifremer, IRD, CNRS, F‐34200, Sète, France

3. Virologie et Immunologie Moléculaires, Institut National de la Recherche Agronomique, Université Paris-Saclay, Jouy-en-Josas, France

4. Univ Polynésie française, Ifremer, IRD, Institut Louis‐Malardé, EIO, F‐98702 Fa’a, Tahiti, Polynésie Française

5. Génomique Métabolique, Genoscope, Institut François Jacob, CEA, CNRS, Univ Evry, Université Paris-Saclay, 91057 Evry, France

6. USR 3278 CRIOBE, EPHE-UPVD-CNRS, Univ. de Perpignan, France

**Corresponding author\***: J. Le Luyer; e-mail:  Jeremy.le.luyer@ifremer.fr

**Keywords:** Microbiome – Gene expression – 16S rRNA – Nanopore – *Tenacibaculum maritimum* – Co-infection

**Supplementary Methods & Results**

## **Metatranscriptomics and metabarcoding sequencing**

**Dual RNA-sequencing.** Mean number of PE raw reads reached 38.68M ± 21.72 sd out of which a mean of 82.47% ± 2.54 remained after filtering (Table S1). After removing reads mapping to non-bacterial domains, number of sequences per individuals reached a mean of 105,642.4 ± 2,186.2 sd.

**Short-reads 16S rRNA metabarcoding sequencing.** We amplified the V4 region for 10–15 replicate individuals per condition. Mean number of PE raw reads reached 231,164.02 ± 36,542.99 sd out of which a mean of 82.47 ± 2.54 remained after filtering (Table S2). After removing reads mapping to non-bacterial domains, the number of sequences per individual reached a mean of 105,642.4 ± 2,186.2 sd. We identified a total of 2577 ASV across the 55 individuals.

**Nanopore full 16S rRNA metabarcoding sequencing.** We amplified the full 16S rRNA for 8 individuals randomly subsampled from the infected group 24 hpi to more precisely and accurately characterise the composition of the main skin microbes. The mean number of SE reads reached 60,019.62 ± 33,778.99 sd after pre-processing (with a minimum of 29,520 sequences).

*P. orbicularis* transcriptome reference

**Sampling, RNA extraction and sequencing.** One individual (50 dph) of *Platax orbicularis* was first sampled for pronephros, tegument, liver and intestine tissue following sampling of two other individuals’ samples for gonads (one male and one female), to build the host transcriptome. Total RNA was extracted from the TRIZOL mix, quantity/integrity and purity were validated by both Nanodrop readings (NanoDrop Technologies Inc.) and Bioanalyzer 2100 (Agilent Technologies). RNA was dried in RNA-stable solution (Thermo Fisher Scientific) following manufacturer’s recommendations and shipped at room temperature to McGill sequencing platform services (Montreal, Canada). TrueSeq v2 kit (Illumina, San 260 Diego, Ca, USA) was used to prepare mRNA depleted libraries that were multiplexed (13-14 samples by lane) and sequenced on HiSeq4000 100 bp PE sequencing device. Individuals that served for the transcriptome assembly were not included in this experiment.

**Transcriptome assembly**. A total of 382.34 M PE 100 bp raw reads (mean 63.72 M ± 7.44 sd) were filtered using Trimmomatic v0.36 [1], with minimum length (36 bp), trailing and leading thresholds of 26 and 26; respectively), implemented in Trinity v2.5.1 [2]. Read quality was assessed with FastQC v0.11.5 (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/). Reads were assembled into transcripts using Trinity v2.5.1 [2] and default parameters. The raw transcriptome was then processed in order to reduce redundancy. First, open-reading frames (ORFs) for each transcript were predicted using ‘*LongOrfs*’ function implemented in Transdecoder v5.3.0 [2, 3]. Only the transcripts containing an ORF of at least 100 amino acids were conserved. Then, only the most expressed isoform for each gene with a minimum mapping rate of 0.5 transcripts per million (TPM) was conserved. Illumina adapters were screened in the transcriptome using a BLASTn (version 2.6.0) approach and adapter list (http://omicsoft.com/downloads/ngs/contamination\_list/v1.txt). Reads were then mapped back onto the filtered transcriptome to evaluate individual mapping rate with BWA mem v0.7.15. For quality checks, the *de novo*transcriptome completeness was assessed with the BUSCO v3.0.2 [4] metazoan single-copy (n = 978) database. We added a final step to look for bacterial contamination by using BLASTn against the NCBI nt database (release 2018-08-27). Transcripts having a hit on Bacteria (e-value<10e-4) were discarded. The resulting transcriptome was then annotated using the Trinotate pipeline v3.1.1 (https://github.com/Trinotate/Trinotate.github.io) following standard guidelines. Detailed procedures for host transcriptome assembly are available in a Github repository (<https://github.com/paulineauffret/Transcriptome_platax>). Transcriptome statistics are given in Table S3.

## **Metatranscriptomic functional analysis**

**Differential expression and Gene Ontology enrichment.** We used a combination of differential expressionand network analyses to explore host and pathogen changes in gene expression profiles during and post infection. The genes were then used for comparing functional differences based on GO enrichment analyses. Details of the results are given in Table S4**.**

References

1. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics. 2014;:btu170.

2. Haas BJ, Papanicolaou A, Yassour M, Grabherr M, Blood PD, Bowden J, et al. De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. Nature Protocols. 2013;8:1494–1512.

3. Celaj A, Markle J, Danska J, Parkinson J. Comparison of assembly algorithms for improving rate of metatranscriptomic functional annotation. Microbiome. 2014;2:39.

4. Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics. 2015;31:3210–2.

**Supplementary Figures and Table s**



Figure S1: Proportion of reads mapped to the microbial compartment. Read origin was dissociated *in silico*. Reads were considered to originate from the microbial compartment when no mapping was apparent in the fish transcriptome. Ctl-24h: *control24h*; Ctl-96h: *control96h*, Inf-24h: *infected24h*; Res-96h: *resistant96h*, groups. Different letters indicate significant differences, *P* < 0.05, Dunn’s test.



Figure S2: Genetic variation and relatedness across established phenotypes. Principal component analysis for the total filtered dataset including 13,448 bi-allelic markers.

****

Figure S3: Alpha-diversity estimates across groups. Alpha diversity was computed using A) Shannon (H’) and B) Fisher indexes. Ctl-24h: *control24h*; Ctl-96h: *control96h*, Inf-24h: *infected24h*; Res-96h: *Resistant96h*, groups. Different letters indicate significant differences, *P* < 0.05, Tukey’s HSD. Each dot represents a single individual.



Figure S4: Taxa enrichment and beta-diversity dissimilarities across groups. A) ASV enrichment between *infected24h* (positive log2FC) and *control24h* (negative values). Colours represent different orders. The y-axis reports shrunken log2FC. Horizontal dash lines represent the log2FC threshold for significance (|log2FC| > 2). Nas represent missing taxonomic information for this ASV; B) ASV enrichment between *resistant96h*(positive log2FC) and *control96h* (negative values). Colours represent different orders. The x-axis represents associated genera; the y-axis reports shrunken log2FC. Nas represent missing taxonomic information for this ASV. C) PcoA based on Bray-Curtis distance values. Ellipses represent 95% CI intervals. Ctl-24h: *control24h*; Ctl-96h: *control96h*, Inf-24h: *infected24h*; Res-96h: *Resistant96h*, groups. Smaller points code for 24 hpi, larger points for 96 hpi groups.



Figure S5: Relative abundance of bacterial genera in the *infected24h* group reported with Nanopore. 16S rRNA sequences and their relative abundance were generated from Nanopore reads (see methods). These sequences were blasted against the NCBI nt database (BLASTn; e-value<10-5). Dot size represents the number of unique sequences per genus.



Figure S6: Total Illumina PE read counts for the most abundant species in the microbial compartment. The selected reference species were the most abundant species represented in the Nanopore 16S rRNA analysis (see methods). T.mar: *T. maritimum*; V.har: *Vibrio harveyi*; A.medit: *Altermonas mediteranea*; P.phe: *Pseudoalteromonas phenolytica*; V.algi: *Vibrio alginolyticus*; and S.yano: *Sphingobium yanoikuyae*. Each dot per species represents one individual.



Figure S7: Plot of expression levels of in vitro and in vivo groups for all virulence-related genes previously identified. The gene list was obtained from whole-genome analysis in *Tenacibaculum* spp. [19]. Asterisks indicate genes with a significant difference between groups (Shrunken |log2FC| > 2; FDR < 0.01).

Table S1: Individual mapping statistics against a combination of the *P. orbicularis* transcriptome and bacterial genomes. Transcriptome column refers to samples included in the metatranscriptome assembly. hpi = hours post-infection. PE = paired end. Mapping rate values are computed based on mapping against the combined reference transcriptome (host + microbiome).



Table S2: Individual sequencing and mapping statistics for MiSeq Illumina reads. hpi = hours post-infection. PE = paired end.



Table S3: *P. orbicularis* transcriptome statistics.



Table S4: List of Fish and *T. maritimum* DEGs and GO term enrichments

Excel file: LeLuyer\_etal\_microbiome.TabS4.xls