Inter-species Transcriptomic Analysis Reveals a Constitutive Adaptation Against Oxidative Stress for the Highly Virulent *Leptospira* Species

Alexandre Giraud-Gatineau [1,†] Garima Ayachit [2,†] Cecilia Nieves [2] Kouessi C. Dagbo [2] Konogan Bourhy,[2] Francisco Pulido [2] Samuel G. Huete [1] Nadia Benaroudj [1] Mathieu Picardeau [1,‡,*] and Frédéric J. Veyrier [2,‡,*]

[1]Microbiology Department, Institut Pasteur, Université Paris Cité, Biology of Spirochetes Unit, Paris, France

[2]INRS-Centre Armand-Frappier Santé Biotechnologie, Bacterial Symbionts Evolution, Laval, Quebec H7V 1B7, Canada

[†]These authors contributed equally.

[‡]These authors contributed equally.

***Corresponding authors:** E-mails: frederic.veyrier@inrs.ca; mathieu.picardeau@pasteur.fr.

Associate editor: Miriam Barlow

Abstract

Transcriptomic analyses across large scales of evolutionary distance have great potential to shed light on regulatory evolution but are complicated by difficulties in establishing orthology and limited availability of accessible software. We introduce here a method and a graphical user interface wrapper, called Annotator-RNAtor, for performing inter-species transcriptomic analysis and studying intragenus evolution. The pipeline uses third-party software to infer homologous genes in various species and highlight differences in the expression of the core-genes. To illustrate the methodology and demonstrate its usefulness, we focus on the emergence of the highly virulent *Leptospira* subclade known as P1+, which includes the causative agents of leptospirosis. Here, we expand on the genomic study through the comparison of transcriptomes between species from P1+ and their related P1- counterparts (low-virulent pathogens). In doing so, we shed light on differentially expressed pathways and focused on describing a specific example of adaptation based on a differential expression of PerRA-controlled genes. We showed that P1+ species exhibit higher expression of the *katE* gene, a well-known virulence determinant in pathogenic *Leptospira* species correlated with greater tolerance to peroxide. Switching PerRA alleles between P1+ and P1- species demonstrated that the lower repression of *katE* and greater tolerance to peroxide in P1+ species was solely controlled by PerRA and partly caused by a PerRA amino-acid permutation. Overall, these results demonstrate the strategic fit of the methodology and its ability to decipher adaptive transcriptomic changes, not observable by comparative genome analysis, that may have been implicated in the emergence of these pathogens.

***Key words:* *Leptospira*, pathogen, inter-species, transcriptomic, oxidative stress.**

Introduction

Bacteria are the most diverse and abundant cellular organisms on Earth, and in recent years, genomics has shed light on this diversity. The evolution of these species is occurring through mutations, deletions, and horizontal gene transfers which contribute to overall speciation events and adaptation (Arnold et al. 2022). One of the possible evolutionary trajectories includes establishing symbiotic interactions with a mammalian host. It has been shown that the vast majority of host-associated symbionts evolved from free-living environmental counterpart (Sachs et al. 2011). The interplay between initially unrelated aspects of the biology of microorganisms and their new host can lead to a closer host–microbe symbiosis than can later evolve within the parasite-mutualist continuum (Drew et al. 2021; Sieber et al. 2021). Understanding the genesis of the bacteria–host interaction is crucial for deciphering the causes of bacterial pathogenesis.

With advances in NGS sequencing technologies, it has become easier to infer and mine information from genomic data. Apart from this, studies have also highlighted the importance of transcriptomic comparison in discerning the mechanisms of regulatory evolution in several bacteria (Said-Salim et al. 2006; Perez and Groisman 2009; Bryant et al. 2021). Nevertheless, these studies are often limited to comparisons between bacteria from the same species (e.g. comparison of *Escherichia coli* strains or comparison

Received: October 20, 2023. **Revised:** February 29, 2024. **Accepted:** March 07, 2024

of *Mycobacterium tuberculosis* complex strains). Despite the widespread use of RNAseq as a method to study gene expression changes and the existence of several tools for investigating differential expression of genes between two groups, there are limited available tools or pipelines for studying differential expression in multiple species simultaneously as orthology needs to be first assigned. Indeed, such analyses currently require customized bioinformatic pipelines beyond the reach of all investigators. It is therefore challenging to compare more diverse bacteria to investigate ancestral events, such as those linked to the emergence of clades or subclades of bacteria within a genus (herein referred to as intragenus evolution). Despite these limitations, a recent study has successfully demonstrated the value of comparing the expression of genes between bacteria of the same family that differ in their cell shape. That analysis has revealed the effect of the loss of the division and cell wall (*dcw*) cluster regulator MraZ in multicellular and longitudinally dividing *Neisseriaceae* MuLDi (Nyongesa et al. 2022a). Similarly, another study reported the ancestral nitrogen-fixing root nodule symbionts transcriptome by combining transcriptomics and phylogenomics of multiple species (Libourel et al. 2023). Thus, despite the promise of comparative transcriptomics to elucidate intragenus evolution, an easy to use, accessible informatics tool is lacking.

In the present study, we introduce a method for studying intragenus transcriptomic evolution which is also available within a graphical user interface (GUI) wrapper called Annotator-RNAtor. This bioinformatics pipeline can infer homologous genes in various species and highlight differences in the expression of the core-genes. To illustrate the methodology and demonstrate its usefulness in deciphering intragenus evolution, we focus on the emergence of the highly virulent *Leptospira* subclade, known as P1+ (sometimes referred to as P1hv for highly virulent pathogens). This subclade includes the multiple causative agents of leptospirosis, an emerging zoonotic disease transmitted to humans through exposure to soil or water contaminated with the urine of animal reservoirs. Annually, an estimated 1 million cases of leptospirosis and nearly 60,000 deaths occur, resulting in a loss of 2.9 million disability-adjusted life years (Costa et al. 2015; Torgerson et al. 2015). Recent elaborated phylogenetic analyses have resulted in a comprehensive new genome-based classification scheme of the *Leptospira* genus (Vincent et al. 2019). Interestingly, this study described the genomic features of the specific group of *Leptospira* P1+, which emerged after a specific node (node 1 in Fig. 1) and that is most often associated with severe infections in humans and animals (Thibeaux et al. 2018; Vincent et al. 2019). Here, we demonstrate the utility of the Annotator-RNAtor pipeline and expand on previous genomic studies of this clade (Vincent et al. 2019) through the comparison of transcriptomes between species from the P1+ lineage and their related P1- counterparts (or P1lv for low-virulent pathogens in animal models, see Fig. 1). In doing so, we shed light on several pathways differentially expressed and specific disparity in the regulation of several PerRA-regulated genes in P1+ species. PerRA is a transcriptional regulator of the Fur family that regulates the response to oxidative stress in *Leptospira* (Lo et al. 2010; Zavala-Alvarado et al. 2020). P1+ species exhibit higher expression of the PerRA-controlled catalase encoding gene (*katE*), a well-known leptospiral virulence determinant (Eshghi et al. 2012), correlating with a greater catalase activity and ability to withstand peroxide. Through a thorough characterization of these pathways, we therefore demonstrated the utility of Annotator-RNAtor-based comparative transcriptomics approach from the initial RNAseq to the experimental validation of its findings.

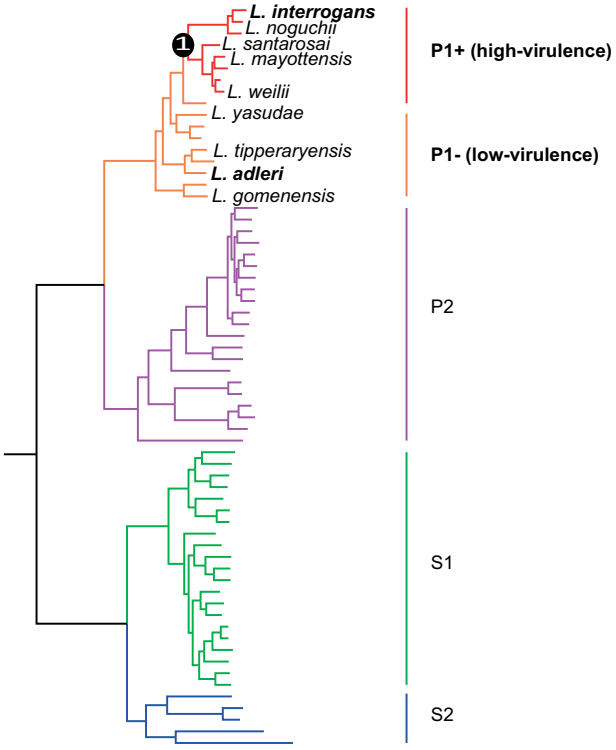

Fig. 1. The P1+ and P1- groups from the P1 subclade in the *Leptospira* genus. The subclade P1, formerly referred to as the "pathogens" lineage, can be separated into two distinct groups: P1+ and P1-. P1+ consists of species associated with severe infections and diverged after a specific node (node 1), while P1- comprises species that have not been isolated from patients and are considered as "low-virulent pathogens". The complete phylogeny (and the associated methods to construct it) has been described previously (Vincent et al. 2019). Species used for RNAseq are indicated whereas organism used for experimental validation are highlighted in bold.

Methods

Bacterial Strains and Culture Conditions

The *Leptospira* strains used in this study, listed in Table 1 (RNAseq) and supplementary table S1, Supplementary Material online (PerR pathway in vitro characterization), were cultivated aerobically in Ellinghausen–McCullough–Johnson–Harris (EMJH) medium at 30 °C with shaking at 100 rpm. The *E. coli* strains Π1 and β2163 were cultivated at 37 °C with shaking, in Luria–Bertani medium supplemented with 0.3 mM thymidine or diaminopimelic acid, respectively.

Table 1 Details of *Leptospira* species used for testing Annotator-RNAtor

Species	Serogroup	Strain	Genome size (Mb)	Group	Transcriptome reads (in million) PE100	Host/source
L. interrogans	Pyrogenes	UP-MMC-NIID-LP	4.7	P1+	4.43	Human/Philippines
L. mayottensis	Mini	200901116	4.3	P1+	4.72	Human/Mayotte
L. noguchii	Panama	201102933	4.8	P1+	4.13	Human/Guadeloupe
L. santarosai	Shermani	LT 821	4.1	P1+	4.20	Rat/Panama
L. adleri	Unknown	201601302	4.9	P1-	4.99	Soil/Mayotte
L. gomenensis	Unknown	KG8-B22	4.3	P1-	3.49	Soil/New Caledonia
L. tipperaryensis	Unknown	GWTS#1	4.7	P1-	4.05	Shrew/Ireland
L. weilii	Celledoni	14535	4.5	P1-	5.32	Human/Laos
L. yasudae	unknown	M12A	4.4	P1-	8.45	Water/Mayotte

RNA Sequencing and Comparisons

To assess the performance of the pipeline, a comparative analysis was conducted on nine genomes of the genus *Leptospira* (Fig. 1 and Table 1). For all experiments, *Leptospira* species listed in Table 1 were used during the exponential phase of growth. First, exponential phase cultures were resuspended in TRIzol lysis reagent (ThermoFisher Scientific). Nucleic acids were extracted with chloroform and precipitated with isopropanol as previously described (Zavala-Alvarado and Benaroudj 2020) and DNA was removed by DNAse treatment using the Turbo DNA-free kit (Thermo Scientific). Prior to cDNA synthesis (RevertAid RT Reverse Transcription Kit, K1691, Thermo Scientific), rRNA was removed from 1 µg of total RNA using the NEBNext rRNA Depletion Kit. The sequencing was performed at the Genome Québec Innovation Centre (McGill University, Montréal, Canada). Briefly, sequencing libraries were constructed using the Illumina Stranded mRNA Prep Kit according to the manufacturer's instructions. Subsequently, 100 bp pair-end sequencing was carried out using the NovaSeq 6000 System. The FastQ reads have been deposited in the SRA database PRJNA998607.

Annotator-RNAtor Software for RNA Comparisons

General Description

There are two primary modules to analyze data in this pipeline (Fig. 2). In the initial steps, the Annotator module assigns homologous sequences in the different species studied and reannotates the genomes with common user-defined unique tags. A gene association table is generated at the end of this module which shows presence–absence of genes in the genomes. Once the gene association table is generated, the RNAtor module is used to map transcriptome data onto the genomes to generate a counts matrix file (number of reads by genes). The final counts comparison file shows the expression of the homologous genes in all the species studied using homogenized tags. Species-specific gene counts can also be seen.

Overall Analyses

Comparative expression analyses for *Leptospira* species were conducted as follows. The genome sequences of the nine evaluated *Leptospira* species (Table 1) were reannotated using Prokka v1.14.5 to standardize annotations. In this experiment, we used RNA extraction from five species that are considered as replicates of P1+ species and four for P1. For each pair-wise genome comparison, a standalone BLASTP search was performed using the respective protein sequences (.faa) files. Network connections were established using a python programming package NetworkX version 2.6.2 (Hagberg et al. 2018), with a similarity threshold set at 60%. Essentially, all proteins showing more than 60% similarity with one of the members (putative homologues) were clustered together. Each cluster of proteins was assigned a name (e.g. Lsp_1), which was then used to replace the original locus tags in the .gff file generated by Prokka. Singletons are not reannotated using the pipeline and can be seen as locus tags in the table. Protein files were also screened for putative homologues using GET_HOMOLOGUES (Contreras-Moreira and Vinuesa 2013) through OrthoMCL method. The reannotated .gtf files were utilized to map the reads to their corresponding genomes using BWA.

Input Data and Formats Using the GUI

The pipeline accepts whole-genome sequence FASTA files. If the genomes are already annotated, it requires genome FASTA, protein FASTA, genome feature files, and coding sequences for Annotator. The RNAtor module requires related FASTQ files as input. Currently, the pipeline supports analysis of both paired-end short-read as well as long-read nanopore data.

Third Party Software and Dependencies, Data Processing and Clustering

The pipeline utilizes Prokka for annotating the genomes, followed by BLAST+ programs (Camacho et al. 2009; Seemann 2014). NetworkX (Hagberg et al. 2018) and GET_HOMOLOGUES v16052022 (Contreras-Moreira and Vinuesa 2013) are used for screening homologues in subsequent steps (Contreras-Moreira and Vinuesa 2013). Homologues can be screened based on the percent similarity/identity as parameters. Additionally, GET_HOMOLOGUES provides options to cluster homologues using bidirectional best-hit, COGtriangles v2.1 (Kristensen et al. 2010), or OrthoMCL v1.4 (Li et al. 2003). At the end of the process, a tab-separated file is generated, in which Annotator-RNAtor creates a presence (locus) or absence (—) map, facilitating the finalization of the analyses. The RNAtor menu in Annotator-RNAtor implements

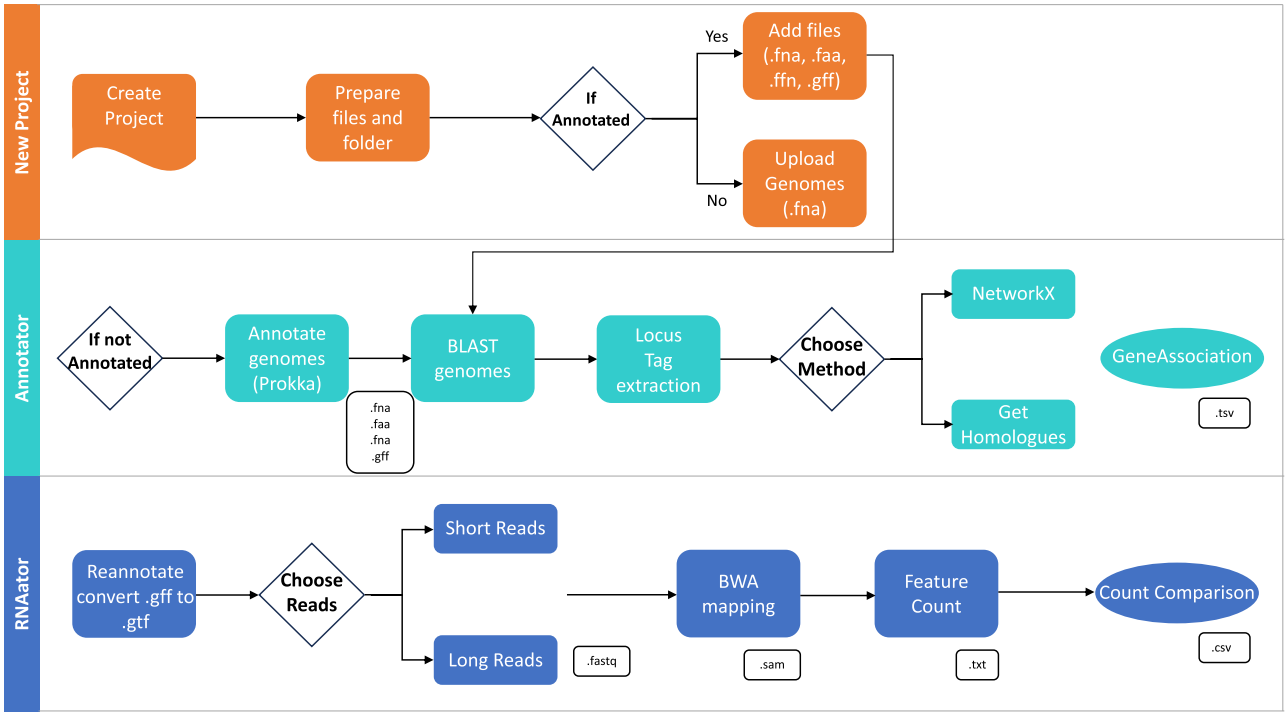

Fig. 2. Pipeline workflow and modules.

the bwa (Li and Durbin 2009) and bwa-mem commands. Annotator-RNAtor is designed to utilize the annotation files from NCBI. RNAtor invokes Pandas and FeatureCounts of Subread (Liao et al. 2014).

Supported Platforms and Availability

Annotator-RNAtor is built using Python3 with PyQt bindings for a user-friendly GUI interface. The efficiency of Annotator-RNAtor relies on running tasks in parallel, utilizing multiple processors with GNU Parallel. The pipeline has been tested on Linux/Ubuntu with 8 GB RAM and 64-bit processors. It includes a user manual with a detailed hands-on tutorial and an installation script that verifies dependencies. Alternatively, due to constant version upgrades in software, the package also includes a conda recipe file to install required dependencies in conda environment. The stand-alone pipeline can be downloaded from GITHUB via https://github.com/BactSymEvol/Annotator.

Statistical Analyses

In the RNAtor module, the .gtf and .sam files were utilized to perform read counts using featureCounts v2.0.1 from the Subread package. The count files for each sample were combined into a table using a custom script, and these results were subsequently analyzed using DESeq2 version 3.14 (Love et al. 2014). The different species, with RNA extracted once, have been considered as "biological replicates" for statistical purposes. To retain only genes that are conserved between the species (soft-core genome of our dataset) and exclude genes unique to one or few species (from shell and cloud genomes of our dataset), genes without any counts in less than eight species were not considered. The genomes were grouped into P1+ and P1- categories. The data were normalized and transformed using the VarianceStabilizingTransformation (VST) function to yield a homoscedastic matrix of values. Further, Log2 fold change values were extracted with an FDR-adjusted *P* values (padj) cut-off of 0.05. To facilitate the interpretation of results, the differentially regulated genes are presented using the locus tag of *Leptospira interrogans* serovar Manilae strain UP-MMC-NIID-LP. We defined two datasets of differentially expressed genes, over and underexpressed in P1+ vs. P1- species based on a fold change threshold of >2 at padj < 0.05. We then investigated the biological functions of the genes in each dataset and the putative pathways that linking them through a Clusters of Orthologous Genes (COG) analyses (Tatusov et al. 2000). The protein-coding sequences from both datasets were also classified based on COG using eggNOG mapper (options –e-value 0.001 –score 60 –pident 40 –query_cover 20 –subject_cover 20 –target_orthologs all –pfam_realign denovo) (Huerta-Cepas et al. 2017; Cantalapiedra et al. 2021). The representativeness of COG categories in each dataset was calculated by determining the ratio of protein-coding genes in each category, normalized by the total number of protein-coding genes in *L. interrogans* serovar Manilae (3,572). Gene ontology (GO) enrichment analyses were performed using the Cytoscape app ClueGO (version 2.5.3) (Bindea et al. 2009). The following parameters were used: only pathways with pV ≤ 0.05, minimum GO level $= 3$, maximum GO level $= 8$, Min GO family > 1, minimum number of genes associated to GO term $= 3$, and minimum percentage of genes associated to GO term $= 4$. Enrichment *P*-values were calculated

using a hypergeometric test (P-value < 0.05, Bonferroni corrected).

Quantitative Reverse Transcription PCR (RT-qPCR)

Total RNA from *Leptospira* strains was extracted using QIAzol lysis reagent (Qiagen) and purified with RNeasy columns (Qiagen). Reverse transcription of mRNA to cDNA was carried out using the iScript cDNA Synthesis kit (Bio-Rad), followed by cDNA amplification using the SsoFast EvaGreen Supermix (Bio-Rad). All primers used in this study are listed in supplementary table S2, Supplementary Material online. Reactions were performed using the CFX96 real-time PCR detection system (Bio-Rad). The relative gene expression levels were assessed according to the $2^{-\Delta\Delta Ct}$ method using *flaB2* (LIMLP_09410) as reference gene.

Inactivation of *perRA* of *L. adleri*

L. interrogans Δ*perRA* was obtained as previously described (Lo et al. 2010). Inactivation of *perRA* gene in *L. adleri* was performed by allelic exchange, replacing the *perRA* coding sequence with a kanamycin resistance cassette. The kanamycin resistance cassette flanked with 0.8 kb sequences homologous to the adjacent *perRA* sequences was obtained by gene synthesis (GeneArt, Life Technologies), and subsequently cloned into an *E. coli* vector unable to replicate in *Leptospira*. The resulting suicide plasmid (pSΔ*perRA*, supplementary table S3, Supplementary Material online) was introduced into *L. adleri* by electroporation using the Gene Pulser Xcell (Bio-Rad) as previously described (Picardeau et al. 2001), and transformants were plated on EMJH supplemented with 50 µg/mL kanamycin. Individual kanamycin-resistant colonies were selected and screened by PCR (using PERADL1and PERADL2 primer set, listed in supplementary table S2, Supplementary Material online) to identify double cross-over events.

Complementation of *Leptospira perRA* Mutants

PerRA expression measurements in all *perRA* mutants were performed with the leptospiral replicative vector pMaOri (Pappas et al. 2015) (supplementary table S3, Supplementary Material online). A vector containing the *L. interrogans perRA* gene under the control of its native promoter (pNB138 named here $perRA_{P1+}$) was previously obtained (Pappas et al. 2015; Kebouchi et al. 2018). The vector harboring the *L. adleri perRA* gene, along with its native promoter region (200 bp upstream region), was amplified from genomic DNA of *L. adleri* (using ComAdPerRA1 and ComAdPerRA2 primer set, supplementary table S2, Supplementary Material online) and cloned between the SacI and XbaI restriction sites in the pMaORI vector. The absence of mutations in the *perRA* locus in the resulting plasmid ($perRA_{P1-}$, supplementary table S3, Supplementary Material online) was confirmed by DNA sequencing (Eurofins). Then, the $perRA_{P1-}$ plasmid was introduced into *L. interrogans* Δ*perRA* and *L. adleri* Δ*perRA*, and the $perRA_{P1+}$ plasmid was introduced into the *L. adleri* Δ*perRA* by conjugation using the *E. coli* β2163 conjugating strain, as previously described (Picardeau 2008). *Leptospira* conjugants were selected on EMJH agar plates containing 50 µg/mL spectinomycin.

Western Blot Analysis

Total extracts of *Leptospira* were obtained by sonication in a lysis buffer containing 25 mM Tris pH 7.5, 100 mM KCl, 2 mM EDTA, 5 mM DTT, and a protease inhibitors cocktail (cOmplete Mini EDTA-free, Roche). A total of 7.5 µg of each lysate were resolved on a 15% SDS–PAGE and transferred onto nitrocellulose membrane. PerRA was detected by immunoblot, as described previously (Kebouchi et al. 2018).

Determination of Bacteria Viability and Catalase Activity

2×10^8 of exponentially growing *Leptospira* species and *L. interrogans* Δ*katE* were incubated in EMJH with or without 1 mM of H_2O_2 for 30 min at 30 °C. Rezasurin (Alamar Blue Assay, ThermoFisher) was added, and bacteria were further incubated for 24 h before measuring absorbance at 570 and 600 nm. Viability was determined based on the ability of cells to reduce rezasurin into resorufin. The percentage of cell viability was calculated as the ratio of rezasurin reduction for bacteria incubated with H_2O_2 to bacteria incubated in the absence of H_2O_2 as described previously (Mouville and Benaroudj 2020). For colony-forming unit (CFU) determination, *Leptospira* (treated and untreated with H_2O_2) were diluted in EMJH and plated on EMJH agar plates. Colonies were counted and the percent survival (% of CFU) was calculated as the ratio of CFU for bacteria incubated with H_2O_2 to bacteria incubated without H_2O_2. A bacterial culture containing about 10^9 exponentially growing *Leptospira* was used to determine the catalase activity using the Catalase Activity Assay Kit (Abcam), according to manufacturer's instructions.

PerRA Alignment Analysis

Homologous sequences of the PerRA protein in *Leptospira* species were searched with BLASTP version 2.13.0 and HMMer version 3.3.1 against the reference database of the 68 *Leptospira* spp (Vincent et al. 2019). Only the sequences with an *e*-value ≤ 0.01 were retained. The results were subsequently aligned by MAFFT version 7.467 using the *L-INS-I* algorithm, and tree inference was performed with IQ-TREE version 2.0.6 with the best-fit model JTTDCMut+F+R5. Putative orthologues of PerRA were extracted from this phylogeny, and the final alignment was refined using MAFFT (version 7.467 using the *L-INS-i* algorithm) with only these sequences. The alignment of PerRA among *Leptospira* spp. was visualized with Jalview software (using Clustal for sequence coloring; Waterhouse et al. 2009). Multiple sequence alignments of all P1+ and P1- species were visualized using *Alvis* software version 0.1. The residues of PerRA involved in the coordination of the regulatory metal were represented using Mol* 3D Viewer of PDB website.

Table 2 Homologous genes (core genome of species in our dataset) identified and reannotated in all P1 *Leptospira* genomes

NetworkX	GET_HOMOLOGUES
2,370	2,323

Mutagenesis of *perRA* Sequence

The H89N substitution of *L. interrogans* perRA ORFs in $perRA_{P1+}$ was conducted using the QuikChange Lightning Multi Site-Directed Mutagenesis Kit (Agilent) according to the manufacturer's protocol and the mutagenic oligonucleotide primer PerRA_c265a. (supplementary table S2, Supplementary Material online). The presence of the c265a mutation and the absence of additional mutation in the *perRA* locus in the obtained plasmid ($perRA_{P1+\text{-}H89N}$) was confirmed by DNA sequencing (Eurofins). Subsequently, the plasmid $perRA_{P1+\text{-}H89N}$ was introduced into *L. interrogans* and *L. adleri* Δ*perRA* mutants by conjugation using the *E. coli* β2163 strain.

In Vivo Animal Studies

Protocols for animal experiments conformed to the guidelines of the Animal Care and Use Committees of the Institut Pasteur (Comité d'Ethique d'Expérimentation Animale CETEA # 2016–0019), agreed by the French Ministry of Agriculture. All animal procedures carried out in this study complied with the European Union legislation for the protection of animals used for scientific purposes (Directive 2010/63/EU).

Male 4-wk-old Syrian Golden hamsters (RjHan:AURA, Janvier Labs) were infected (four per group) by intraperitoneal injection with bacterial suspensions containing 10^6 or 10^8 of *L. interrogans* or *L. adleri*, respectively, as enumerated using a Petroff–Hausser counting chamber. The animals were monitored daily and euthanized by carbon dioxide inhalation upon reaching the predefined endpoint criteria (sign of distress, morbidity). To assess leptospiral load, blood, kidney, and liver were sampled and DNA was extracted with the Tissue or Blood DNA purification kit (Maxwell, Promega). The burden in blood and tissues was determined by qPCR with the Sso Fast EvaGreen Supermix assay (Bio-Rad) using the *flaB2* gene, and the concentration of host DNA was quantified using the *gapdh* gene. *Leptospira* load was expressed as genomic equivalent (GEq) per µg of host DNA.

Quantification and Statistical Analysis

Data are expressed as means ± standard deviations (SD) of at least three independent biological replicates. Statistical analyses were performed with Prism software (GraphPad Software Inc.), using the *t* test as indicated in the figure legends.

Results

Inferring Homology Using NetworkX or GET_HOMOLOGUES

In the current pipeline (Fig. 2), homology analysis is performed using the Annotator module which considers protein similarity searches instead of DNA, making it more sensitive. We used two methods relying on BLAST to search for homologues. NetworkX (Hagberg et al. 2018) uses a graph method to screen out homologues from the BLAST results (based on the user-provided threshold) whereas GET_HOMOLOGUES (Contreras-Moreira and Vinuesa 2013) uses identity to screen them. These approaches both have their benefits depending on the type of data being analyzed. For example, if the purpose of the analysis is to find gene presence/absence in unrelated species or genera the NetworkX method provides more flexibility when searching for similar sequences. It also takes into account fragmented proteins that are not considered as homologues by other tools. This may provide an advantage while working with fragmented assemblies as well. In this study, we analyzed nine *Leptospira* genomes (Fig. 1 and Table 1) along with their transcriptome data. Annotator module was used to extract homologous genes and generate a gene presence/absence table using both NetworkX and OrthoMCL in GET_HOMOLOGUES method. In total we obtained 2,370 core genes (genes shared by all the species of our database) using NetworkX (60% similarity) and 2,323 core genes using GET_HOMOLOGUES (45% identity) (Table 2). The difference comes from the fact that the NetworkX method is based on similarity between sequences and considers fragmented sequences as well whereas GET_HOMOLOGUES is based on identity and has a more stringent approach. GET_HOMOLOGUES does not take into account fragmented or dissimilar sequences (such as those found when comparing divergent species) and annotates them separately (Contreras-Moreira and Vinuesa 2013).

Transcriptomic Comparison Between *Leptospira* Species Using RNAtor

The reannotated genomes by the two methods generated two new .GTF files. These files were used to map the reads to their corresponding genomes using BWA and the number of reads per gene was determined using featureCounts, which produced a .csv table containing the reannotated genes and their respective counts (supplementary table S4, Supplementary Material online). For inferring differential expression between the P1+ and P1- subclades, standard DESeq2 analysis was employed, excluding genes that were absent or not expressed in more than one species. A total of 2,323 and 2,370 instances were analyzed using data from GET_HOMOLOGUES and NetworkX, respectively (supplementary table S4, Supplementary Material online). The volcano plot (fold change in function of the padj values) in Fig. 3 presents the results obtained with NetworkX (Fig. 3a) and GET_HOMOLOGUES (Fig. 3b). Notably, both methods yielded highly similar results, as the majority of over-expressed and under-expressed genes in the P1+ were identified by both methods (see Venn Diagram, Fig. 3c). NetworkX detected approximately 10% more genes overall. The slight variations in the expression values between NetworkX and GET_HOMOLOGUES may

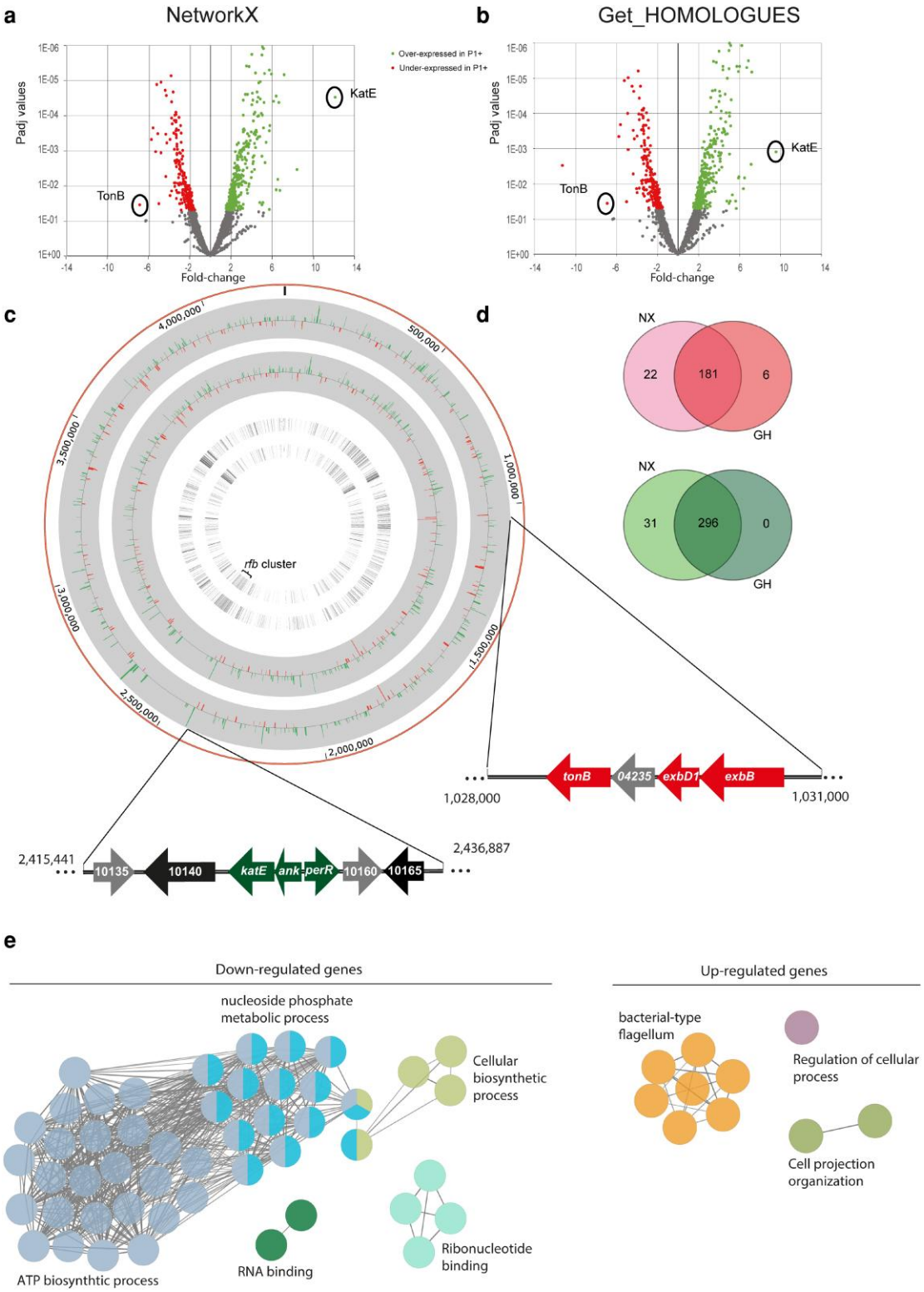

Fig. 3. Comparison of differentially expressed genes in P1+ species compared to P1- using both methods. Volcano plot of RNAseq analysis of P1+ species compared to P1-. Using NetworkX in (a) and GET_HOMOLOGUES in (b). *P*-value is plotted against fold change and were calculated using DeSeq2. Red points represent genes under-expressed (FC < −1.5) and green genes over-expressed (FC > 1.5) in P1+ as compared to P1-. c) Venn diagram showing the comparison of the number of genes over-expressed (red) or under-expressed (green) in P1+ obtained with both methods (NX for NetworkX and GH for GET_HOMOLOGUES). d) Circos plot generated for chromosome 1. The outermost track shows the chromosome, the second shows gene expression values generated using GET_HOMOLOGUES followed by expression values generated using NetworkX. The second innermost track show genes with no homologues in other species determined using GET_HOMOLOGUES and the innermost track using NetworkX. Finally, the *perRA* and *tonB* loci are also indicated with color coded genes (gray: no homologues in other species, black: FC between 1.5 and −1.5, red: FC < −1.5, green: FC > 1.5). e) Gene ontology enrichment analysis of genes whose expression is downregulated or upregulated in P1+ compared to P1-, using Cytoscape app ClueGO.

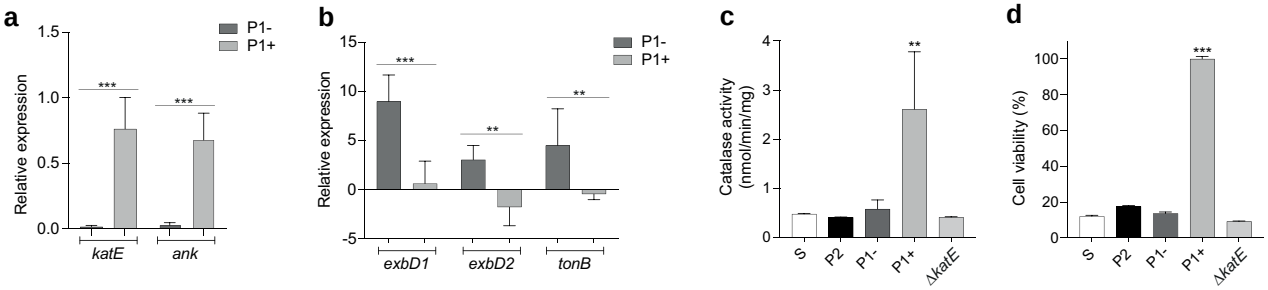

Fig. 4. Higher resistance to oxidative stress in P1+ correlates with higher basal expression of *katE* and catalase activity. a, b) Relative gene expression of *katE* and *ank* (a), *exbD1*, *exbD2*, and *tonB* (b) in P1+ and P1- species measured by RT-qPCR. Relative expression levels were normalized to the *flaB2* gene and compared to *L. interrogans*. c) Measurement of catalase activity of *Leptospira* spp. and *L. interrogans* Δ*katE* mutant. d) Viability of *Leptospira* spp. and *L. interrogans* Δ*katE* mutant exposed to 1 mM of H_2O_2 for 30 min. The percentage of *Leptospira* viability was determined by the ability of bacteria to reduce rezasurin into resorufin with or without H_2O_2. Unpaired two-tailed Student's *t* test was used. $^{**}P < 0.001$, $^{***}P < 0.0001$. One representative experiment (of three) is shown. Error bars represent the mean ± SD. Saprophyte (S): *L. biflexa*; P2 subclade: *L. licerasiae*, *L. fluminis*; P1- subgroup: *L. adleri*, *L. gomenensis*, *L. tipperyarensis*, *L. yasudae*; P1+ subgroup: *L. interrogans*, *L. noguchii*, *L. weilii*, *L. santarosai*, *L. mayottensis*.

be attributed to the separate reannotation approaches and the differences in the parameters utilized (% similarity vs. identity, respectively). Figure 3d presents a circular representation of results obtained for Chromosome 1. The central inner circles represent genes that are not in the soft-core genome. Of note, this representation highlights some nonconserved regions with content variation between species (such as exemplified with the *rfb* cluster of genes in Fig. 3d). As the methodology is intended to compare expression among well-conserved orthologues present in the majority of the species, these genes are therefore excluded from the analyses. The outer circles depict fold change values obtained with both methods, revealing only subtle differences between the two approaches.

Concerning the differentially expressed genes identified using this methodology, a clear overexpression of genes encoding the flagellation machinery (such as *flgE*, *flgK flgB*, *fliH*) was observed in P1+ and an underexpression of general metabolic pathways (such as *atpA* to *atpH* encoding F1F0-ATPase or *cys* pathway implicated in sulfur transport and metabolism) or nucleoside phosphate metabolic process (Fig. 3e). Indeed, the functional COG (Tatusov et al. 2000) analysis revealed a significantly higher representation (20-fold) of the cell motility category in the overexpressed genes compared to the underexpressed ones (supplementary table S5, Supplementary Material online). There is also a 3.5-fold enrichment observed in the category of signal transduction category among the overexpressed genes in comparison to the underexpressed. Conversely, categories involved in the transport and metabolism of amino acids, nucleotides, carbohydrates, lipids, and coenzymes are overrepresented in the underexpressed genes, indicating underutilized metabolic pathways in the P1+ species in vitro.

Additionally, we were also able to observe two genes, *tonB* and *katE*, which are part of two distant loci, that demonstrated strong altered expression between the P1+ and P1- groups (see *ank*/*perR* and *exbD1*/*exbB* genes, respectively, in Fig. 3d). Both loci are part of the PerRA regulon, with *tonB* and *katE* being, respectively, the most underexpressed and overexpressed genes in a *L. interrogans perRA* deletion mutant (Lo et al. 2010; Zavala-Alvarado et al. 2020). Importantly, most of the genes that were differentially expressed in this mutant are also differentially expressed in our analysis of P1+ and P1- expression (supplementary table S6, Supplementary Material online). Most of the genes from this regulon are known to be modulated in the presence of peroxide to protect the cell against oxidative stress. Therefore, our results strongly suggest the hypothesis that an evolutionary significant disparity in the expression of the PerRA regulon exists between the species that form the P1+ and P1- subclades. To both test this hypothesis and its biological implications, as well as to validate our in silico findings, we focused further investigation on this regulon.

A Differential Expression of PerRA-controlled Genes is Associated With an Increased Resistance to Peroxide in P1+

The differential expression of the PerRA-controlled genes (some positively, others negatively) detected with RNAtor between P1+ vs P1- was confirmed by RT-qPCR. Consistent with our transcriptomic data, we observed an increase in the expression of the *ank-katE* operon in P1+ species compared to P1- species (20.2-fold and 47.5-fold increase, respectively, Fig. 4a). Similarly, we confirmed by RT-qPCR that the *tonB* locus (*exbD1*, *exbD2*, and *tonB*), which is known to be positively regulated by PerRA (Lo et al. 2010; Zavala-Alvarado et al. 2020), was under-expressed in P1+ (Fig. 4b).

In *L. interrogans*, the catalase KatE, whose expression is repressed by PerRA, detoxifies H_2O_2 (Eshghi et al. 2012). Inactivation of *katE* resulted in an attenuation of virulence in *L. interrogans*, highlighting its role in pathogenicity (Eshghi et al. 2012). We further confirmed the high basal overexpression of *katE* in P1+ by directly testing the

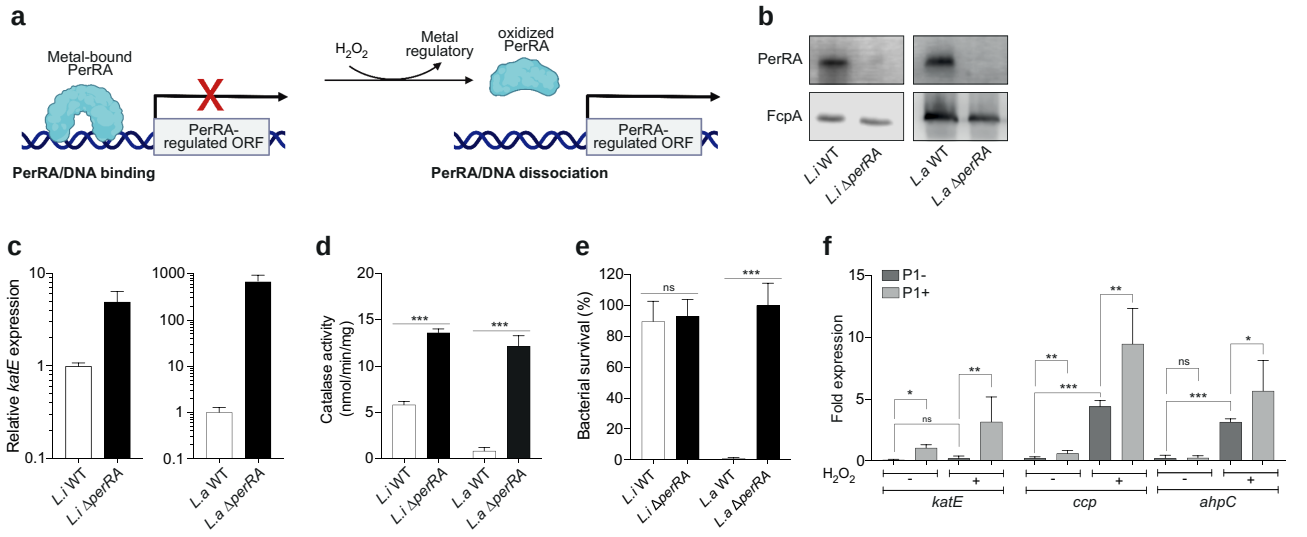

Fig. 5. Difference of PerRA on peroxide-sensing capacity between P1+ and P1- species. a) Schematic representation of the regulation of PerRA binding DNA. In the absence of H_2O_2, PerRA is in a regulatory metal-bound conformation prone to DNA binding, resulting in peroxidase-encoding genes repression. In the presence of peroxide, PerRA is oxidized leading to the release of the regulatory metal and to a conformation switch resulting in dissociation from DNA and repression alleviation. b) PerRA cellular content of the *L. interrogans* (*L.i*) WT, *L.i* Δ*perRA* mutant, *L. adleri* (*L.a*) WT, and *L.a* Δ*perRA* mutant were assessed by immunoblot. An anti-FcpA antibody is used as a control of equal loading. c) Relative expression of *katE* measured by RT-qPCR in *L.i* WT, *L.i* Δ*perRA* mutant, *L.a* WT, and *L.a* Δ*perRA* mutant strains upon exposure to 1 mM of H_2O_2 for 30 min. Relative expression levels were normalized to the *flaB2* gene and compared to *L. interrogans* without H_2O_2 exposure. d) Measurement of catalase activity in total extracts of *L.i* WT, *L.i* Δ*perRA* mutant, *L.a* WT, and *L.a* Δ*perRA* mutant strains. e) Viability of *L.i* WT, *L.i* Δ*perRA* mutant, *L.a* WT, and *L.a* Δ*perRA* mutant strains upon exposure to 1 mM of H_2O_2 for 30 min. Viability was determined by CFU and normalized with H_2O_2-untreated bacteria. f) Relative expression of *katE*, *ccp*, and *ahpC* measured by RT-qPCR in P1+ and P1- species upon exposure to 1 mM H_2O_2 for 30 min. Relative expression levels were normalized to the *flaB2* gene and compared to H_2O_2-untreated *L. interrogans*. Unpaired two-tailed Student's *t* test was used. $^*P < 0.01$, $^{**}P < 0.001$, $^{***}P < 0.0001$, ns: nonsignificant. One representative experiment (of three) is shown. Error bars represent the mean ± SD. P1- subgroup: *L. adleri*, *L. gomenensis*, *L. tipperyarensis*, *L. yasudae*; P1+ subgroup: *L. interrogans*, *L. noguchii*, *L. weilii*, *L. santarosai*, *L. mayottensis*.

catalase activity (Fig. 4c). Interestingly, only P1+ species possess a higher basal catalase activity compared to other *Leptospira* species that contain the *katE* gene. In fact, in P1- species, the catalase activity is comparable to that of a *L. interrogans* Δ*katE* strain. To investigate the impact of a higher *katE* gene expression and catalase activity in P1+ species we analyzed the survival of P1+ and P1- species when exposed to H_2O_2. As expected, the survival of the P1+ species was not impaired by the presence of 1 mM of H_2O_2, whereas the P1- species exhibit only 13.5% survival, a reduced survival comparable to that of the *L. interrogans* Δ*katE* mutant (Fig. 4d). Interestingly, only P1+ species possess a higher basal catalase activity and better tolerance to H_2O_2 compared to other *Leptospira* species, even as compared with species from P and S clades. This suggests that this phenotype specifically evolved during the emergence of P1+ species. These results reveal a distinct advantage for P1+ species in resisting peroxide, despite the genetic relationship and the presence of the *katE* gene in both P1+ and P1- species.

Genes Coding for H_2O_2 Detoxifying Enzymes Are Less Repressed by PerRA in P1+ Than in P1-

We have previously shown that the catalase is repressed by PerRA in *L. interrogans* (Zavala-Alvarado et al. 2020). In the absence of peroxide, PerRA is under the typical metal-bound conformation prone to promoter binding and repression of *katE* gene expression (Kebouchi et al. 2018) (Fig. 5a). Similarly to a canonical PerR, PerRA releases its regulatory metal upon oxidation by H_2O_2, leading to DNA dissociation and alleviation of gene repression. To investigate whether PerRA-controlled gene derepression occurs in a similar manner in P1- species, we generated a Δ*perRA* mutant in *L. adleri*, used here as P1- representative species. After confirming the absence of PerRA production in this mutant by immunoblot (Fig. 5b), we evaluated *katE* expression in the *L. adleri* Δ*perRA* mutant. Similarly to what is observed in *L. interrogans*, *perRA* inactivation in *L. adleri* led to the derepression of *katE* (Fig. 5c) and concomitant increase in catalase activity (Fig. 5d) and in tolerance to peroxide (Fig. 5e). These results indicate that P1- species possess an active PerRA that represses the catalase expression and upon repression alleviation, P1- species exhibit a comparable catalase activity to that of P1+ species. We also examined whether P1- species could sense and respond to H_2O_2 by derepressing *katE*, as previously observed in *L. interrogans* (Zavala-Alvarado et al. 2020). We observed no significant upregulation of the *katE* gene expression in P1- species in the presence of 1 mM of H_2O_2, while P1+ species exhibited a 3.2-fold increase in *katE* expression under these conditions (Fig. 5f). However, the expression of *ahpC* and *ccp*, genes encoding an alkyl hydroperoxide reductase (AhpC) and a cytochrome C peroxidase (CCP), respectively, known for their involvement in

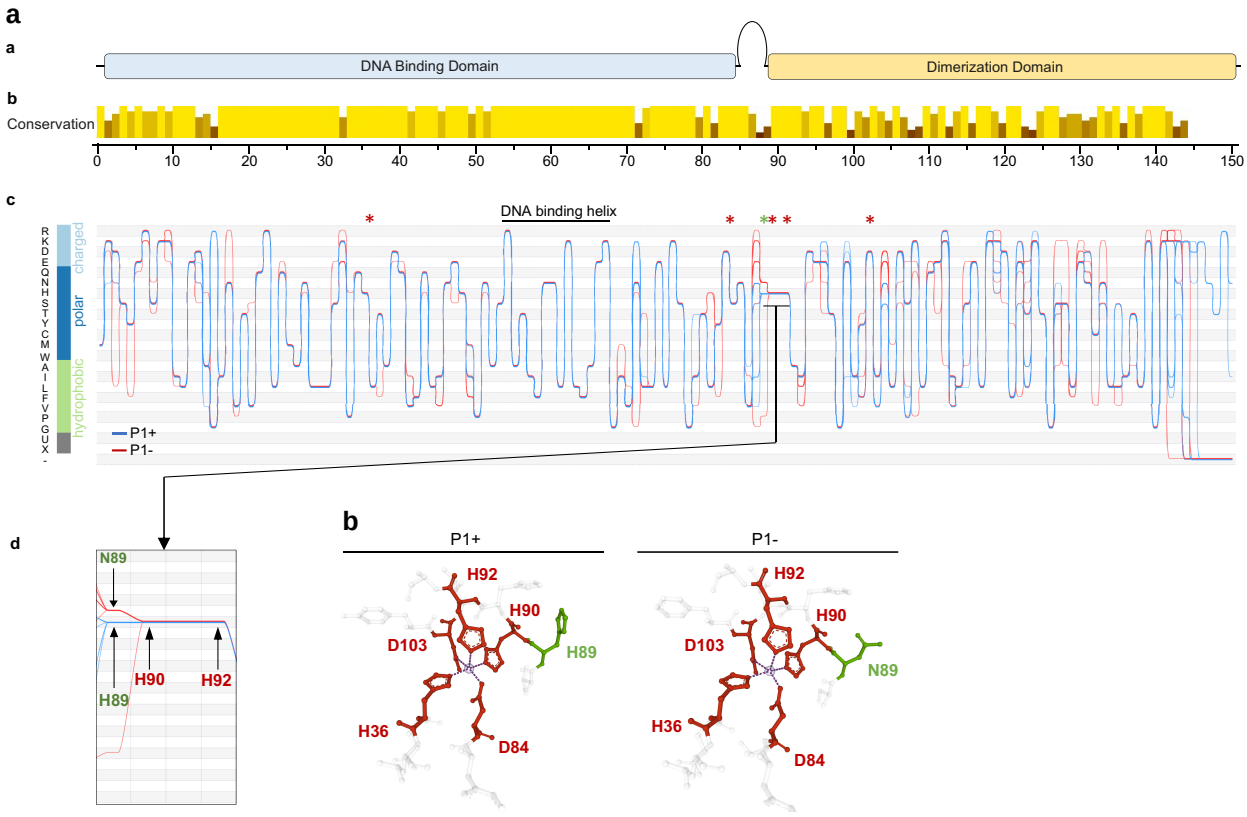

Fig. 6. PerRA protein sequence divergence between P1+ and P1- species. a) PerRA protein sequence alignment of P1+ and P1- species were performed. Schematic representation of PerRA protein sequence (A) is shown with associated domains following by the conservation of each amino acid for P1 species (B). Sequence bundles visualization of PerRA for P1+ and P1- species using Alvis software (C). Bundles plots show P1+ and P1- sequence groups in blue and red lines, respectively, represented against a *y*-axis classified by amino acid hydrophobicity. Each line corresponds to one species. The curved trajectories of lines expose the conservation of residuals by converging to identical positions. Amino acids involved in regulatory metal coordination are indicated by a red asterisk and the amino acid in position 89 is indicated by a green asterisk. PerRA protein sequence alignment between the positions 89 to 92 is zoomed in (D). b) Detailed view of the regulatory metal binding site of PerRA of P1+ species (left) and P1- species (right). The coordination residues (H36, D84, H90, H92, and D103) are labeled in red and the metal in the regulatory metal-binding sites are represented by a sphere. Ligand coordination is symbolized by purple dashed lines. The amino acid position 89 of PerRA is labeled in green.

defense against peroxide in *L. interrogans* (Zavala-Alvarado et al. 2020) was upregulated upon exposure to H_2O_2 in both P1- and P1+ species. Notably, the H_2O_2-triggered derepression was consistently higher in P1+ species (Fig. 5f). Our data show that, while PerRA can sense and respond to H_2O_2 by alleviating gene repression in both P1- and P1+ species, H_2O_2-triggered derepression is greater in P1+ than in P1- species for the same H_2O_2 concentration. This suggests that either $PerRA_{P1+}$ and $PerRA_{P1-}$ have different peroxide-sensing capacity, $PerRA_{P1+}$ being more sensitive to oxidation by peroxide, and/or that $PerRA_{P1+}$ has a lower affinity for DNA than $PerRA_{P1-}$.

Effect of a Residue 89 Substitution in PerRA on the H_2O_2-triggered Derepression

In order to identify amino acid residues responsible for the weaker PerRA-mediated repression in P1+ species, we compared the protein sequences of PerRA from all P1 species (supplementary fig S1A, Supplementary Material online, Fig. 6a). The DNA binding helix as well as the regulatory metal coordination site (H36, D84, H90, H92, D103), are relatively well conserved between P1+ and P1- species. We then searched for PerRA residues conserved among P1+ species but not found in all P1- species. We identified residue 89 as having a strikingly different distribution in P1+ and P1- species. This residue is a histidine in all P1+ species, while it is an asparagine in all P1- species, (except for *L. ainazelensis* which has a proline at this position). We also observed an absence of histidine at the position 89 for the P2 and S species, which contain the *perRA* gene (except for *L. noumeaensis*). This residue is located in a loop connecting two domains in proximity to the regulatory metal coordination site of PerRA (Fig. 6a and b), suggesting that it can participate in H_2O_2 sensing by PerRA. Structure predictions indicate that the presence of an Asn in position 89 does not affect the regulatory metal coordination (Fig. 6b). However, prediction of the structural disordered regions in PerRA suggests that the presence of the His89 of P1+ species results in a lower probability of disordered region compared to the presence of an Asn in position 89 of $PerRA_{P1-}$ (and other species;

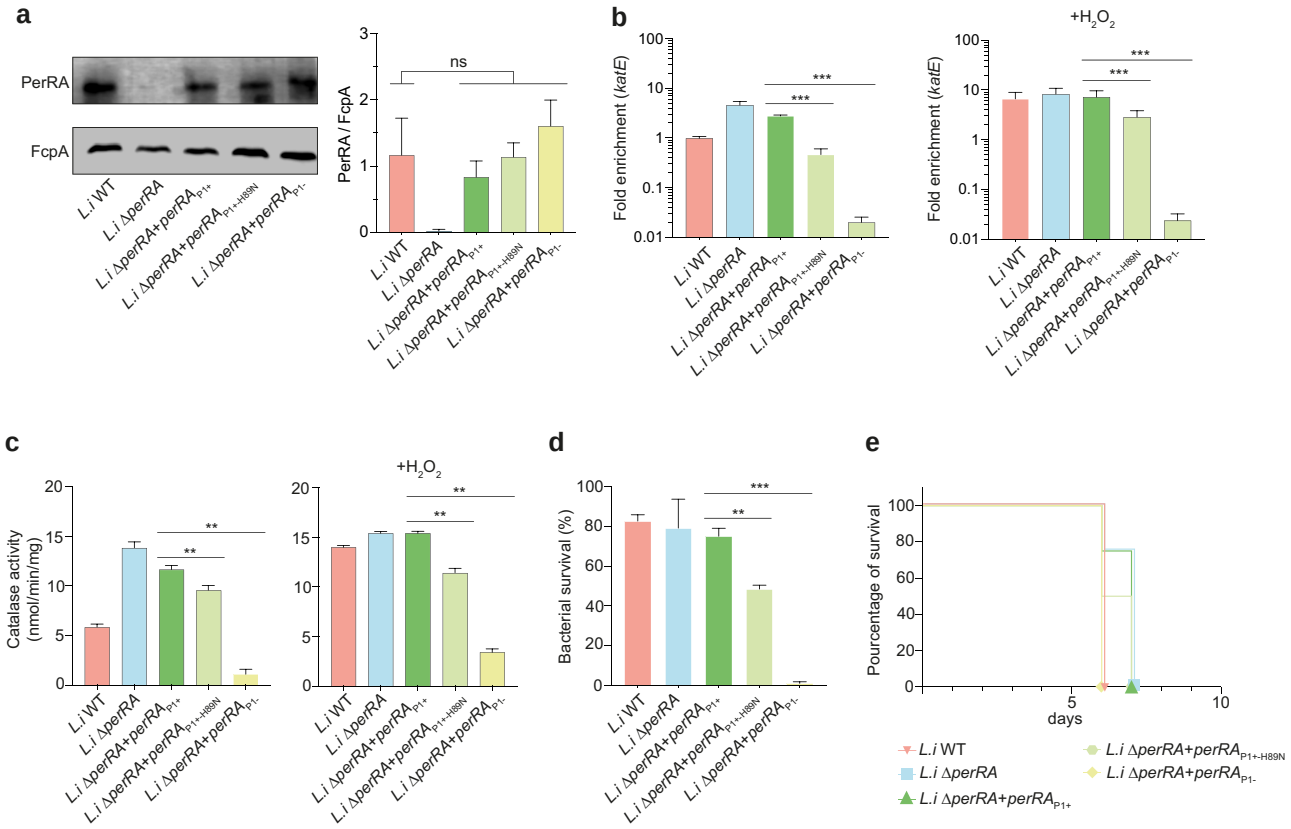

Fig. 7. PerRA from P1+ species is associated with increased peroxide stress tolerance. a) PerRA cellular content of the *L. interrogans* (*L.i*) WT and Δ*perRA* mutant complemented with the indicated plasmids was assessed by immunoblot using an anti-FcpA antibody as a control of equal loading (left panel). Densitometric quantification of PerRA was performed and normalized by FcpA amount (right panel) in *L.i* constructs. *L.i* WT and *L.i* Δ*perRA* contain the empty pMaORI-expressing vector; *L.i* Δ*perRA*+*perRA*$_{P1+}$ contains the pMaORI vector bearing the *perRA* ORF of *L.i*; *L.i* Δ*perRA*+*perRA*$_{P1+-H89N}$ contains the pMaORI vector bearing the *perRA* ORF of *L.i* with the single mutation H89N; *L.i* Δ*perRA* +*perRA*$_{P1-}$ contains the pMaORI vector bearing the *perRA* ORF of *L. adleri*. Data are the means and standard deviation of two independent biological replicates. b) Relative expression of *katE* measured by RT-qPCR in the different *L. interrogans* (*L.i*) strains in the absence (left panel) or presence (right panel) of 1 mM of H_2O_2 (30 min exposure). Relative expression levels were normalized to the *flaB2* gene and compared to *L. interrogans* WT (*L.i*). c) Measurement of catalase activity in total extracts of the different *L. interrogans* (*L.i*) strains in absence (left panel) or in presence (right panel) of 1 mM of H_2O_2 (30 min exposure). d) Survival of the different *L. interrogans* (*L.i*) strains upon exposure to 1 mM of H_2O_2 for 30 min. The number of bacteria was enumerated by CFU and compared to H_2O_2-untreated *Leptospira*. e) The virulence of the different *L. interrogans* (*L.i*) strains was assessed by infecting hamsters ($n = 4$) by peritoneal route with 10^6 leptospires for each construct. For experiments in (a) to (d), one representative experiment of three independent biological replicates is shown. Error bars represent the mean ± SD. Unpaired two-tailed Student's *t* test was used. $^{**}P < 0.001$, $^{***}P < 0.0001$, ns: nonsignificant.

supplementary fig. S1B, Supplementary Material online). Thus, *in silico* analysis suggests that residue 89 could have a role on PerRA differential activity.

To further investigate the role of residue 89 of PerRA, we complemented *L. interrogans* Δ*perRA* with the heterospecific *perRA*$_{P1-}$, with the native *perRA* from *L. interrogans* (*perRA*$_{P1+}$) as well as with the one containing the mutation H89N (*perRA*$_{P1+-H89N}$). We first verified that all complemented strains restored PerRA production to a similar level as *L. interrogans* WT (Fig. 7a). Complementing the *L. interrogans* Δ*perRA* mutant with *perRA*$_{P1+}$, led to a lower *katE* repression than with *perRA*$_{P1+-H89N}$ or *perRA*$_{P1-}$ (Fig. 7b). Interestingly, not only *katE* was dramatically more repressed by *perRA*$_{P1-}$ but repression was not alleviated in the presence of 1 mM of H_2O_2. Complementing the *L. interrogans* Δ*perRA* mutant with *perRA*$_{P1-}$ resulted in 8-fold reduction of the catalase activity and 70-fold greater survival in the presence of H_2O_2 than when complementing the *L. interrogans* Δ*perRA* mutant with *perRA*$_{P1+}$ (Fig. 7c and d). Therefore, differences in *katE* expression correlated with the catalase activity, in the presence and absence of 1 mM of H_2O_2 and with the ability of the strains to tolerate H_2O_2. Overall, our findings demonstrate that PerRA$_{P1+}$ represses *katE* to a lesser extent than PerRA$_{P1+-H89N}$ or PerRA$_{P1-}$. It is important to note that complementation with *perRA*$_{P1-}$ consistently led to a more dramatic effect than when complementing with *perRA*$_{P1+-H89N}$, indicating that permutation of the residue 89 in PerRA$_{P1+}$ could not solely recapitulate the extent of *katE* repression exerted by PerRA$_{P1}$. Although we observed a decrease in peroxide stress survival for *L. interrogans* Δ*perRA* expressing *perRA*$_{P1+-H89N}$ or *perRA*$_{P1-}$, these strains did not show attenuation of virulence in hamsters (Fig. 7e).

To further assess the importance of PerRA and the H89 residue in mediating a lesser *katE* repression in P1+ species, we also complemented the *L. adleri* Δ*perRA* mutant with *L. interrogans perRA* ($perRA_{P1+}$), *L. interrogans perRA* with mutation H89N ($perRA_{P1+\text{-}H89N}$) or *perRA* from *L. adleri* ($perRA_{P1-}$; Fig. 8a). Consistent with results obtained in *L. interrogans*, complementing the *L. adleri* Δ*perRA* with $perRA_{P1+}$ resulted in a lower repression of *katE* than when this mutant was complemented with $perRA_{P1-}$ or with $perRA_{P1+\text{-}H89N}$ (Fig. 8b). Furthermore, the reduced repression of *katE* observed in the mutant *L. adleri* Δ*perRA* complemented with $perRA_{P1+}$ correlated with higher catalase activity in the presence of H_2O_2 and an increase in tolerance to peroxide (Fig. 8c and d). Since P1- species, including *L. adleri*, are rapidly cleared from hamsters, we could solely assess *Leptospira* burden in blood, kidney, and liver 1 d post-infection but no significant difference was observed between the different *L. adleri* strains (Fig. 8e). This indicates that the difference of catalase activity species is not the only determinant of the difference observed in this model of acute infection between P1+ and P1.

Altogether, these findings demonstrate that difference in *katE* repression and ability to tolerate H_2O_2 between the P1 subgroups is mediated only by PerRA and that the residue 89 participates in tuning this repression.

Discussion

Differential gene expression studies across multiple species require consideration of processes for normalization, bias reduction, and enrichment. Critically, analysis of gene expression in multiple species is highly dependent on a robust phylogeny and identification of orthologous genes. In their absence, false predictions may result, grouping nonrelated genes into the same clusters (Kristiansson et al. 2013). These analyses are further complicated by lack of an accessible, easy-to-use bioinformatics pipeline.

In this study, we developed a methodology and a stand-alone GUI wrapper to analyze intra-genus species using homology. The pipeline is particularly useful for understanding subtle variations in the transcriptomic repertoire of phylogenetically related species within the same genus, which may exhibit varied phenotypes such as morphological features or pathogenicity. However, this approach has certain limitations: orthology may be incorrectly assigned in poorly annotated or divergent genomes, which could lead to bias in the analysis. To mitigate this risk, we used standardized annotation and employ two methods for homology assignment. In addition, even though they concern only few genes from the core genome, duplications of genes in specific species also represent a significant challenge for analysis. Nevertheless, we demonstrated that this tool is able to identify major changes in core genes expression between two groups of species.

Herein, we show that there is a transcriptomic change associated with the emergence of P1+ species. Functional analysis of COG indicates a significant presence of genes encoding hypothetical proteins in both overexpressed and underexpressed genes in P1+ compared to P1-. As previously mentioned, the overexpressed genes in P1+ include an overrepresentation of those related to motility, a key factor in pathogenicity, as well as genes associated with signal transduction, among others. Conversely, the underexpressed genes in P1+ include those involved in metabolism and transport of amino acids, carbohydrates, lipids, and genes related to energy production. Previous studies have shown a prevalence of genes encoding energy metabolism and transport-related pathways in free-living leptospires (Fouts et al. 2016). Given that P1- are predominantly isolated from environmental samples, the underexpression of these pathways in P1+ may reflect specific adaptations associated with the transition from a primarily environmental diverse lifestyle to a more restricted host-associated one.

The detection of transcriptomic changes needs to be complemented by studies that test and explain the mechanisms of adaptation. Previously, we have explained transcriptomic changes by the loss of a regulator (Nyongesa et al. 2022b). Here, we were able to demonstrate the major impact of subtle adjustments in the protein sequence of the regulator PerRA. Of note, we already designed a tool able to pinpoint these amino acid changes and selection for the exploration of intragenus evolution (Guerra Maldonado et al. 2020). A single amino acid residue change in a regulator like PerRA has a major impact. In P1+ species, it leads to a higher basal *katE* expression, catalase activity, and tolerance to H_2O_2. Residue 89 is located within a loop at the hinge of the two domains in PerRA, in the vicinity of the regulatory metal coordination site (H36, D84, H90, H92, D103). The nature of this residue could impact the flexibility of this loop and thereby influence the dynamism of the metal and peroxide-triggered allosteric conformational switch. In that case, the presence of a histidine at position 89 in $PerRA_{P1+}$ would result in a higher amount of PerRA in the metal-free conformation with low affinity for promoter region. By analogy with PerR from *Bacillus subtilis* (Lee and Helmann 2006), H36 and H90 are the two PerRA residues whose oxidation by H_2O_2 leads to the release of the regulatory metal and dissociation from DNA. The residue at position 89 is in close proximity to the oxidized residues, therefore, it can also be speculated that the histidine 89 in $PerRA_{P1+}$ is also a site of oxidation and, even if it does not directly participate in the regulatory metal binding, its oxidation could result in destabilizing the metal coordination. The asparagine side chain is less prone to oxidation than that of histidine, therefore the presence of an asparagine at the corresponding position in $PerRA_{P1-}$ would lower the chance of oxidation. Other residues of PerRA might also participate in the metal and peroxide-triggered conformational switch, thereby fine-tuning gene expression. In fact, heterologous complementation experiments show that *katE* repression was greater when expressing $perRA_{P1-}$ than when expressing $perRA_{P1+\text{-}H89N}$ in the *L. interrogans* Δ*perRA* mutant. Therefore, although important, the H89N alone could not explain the higher gene repression by PerRA in

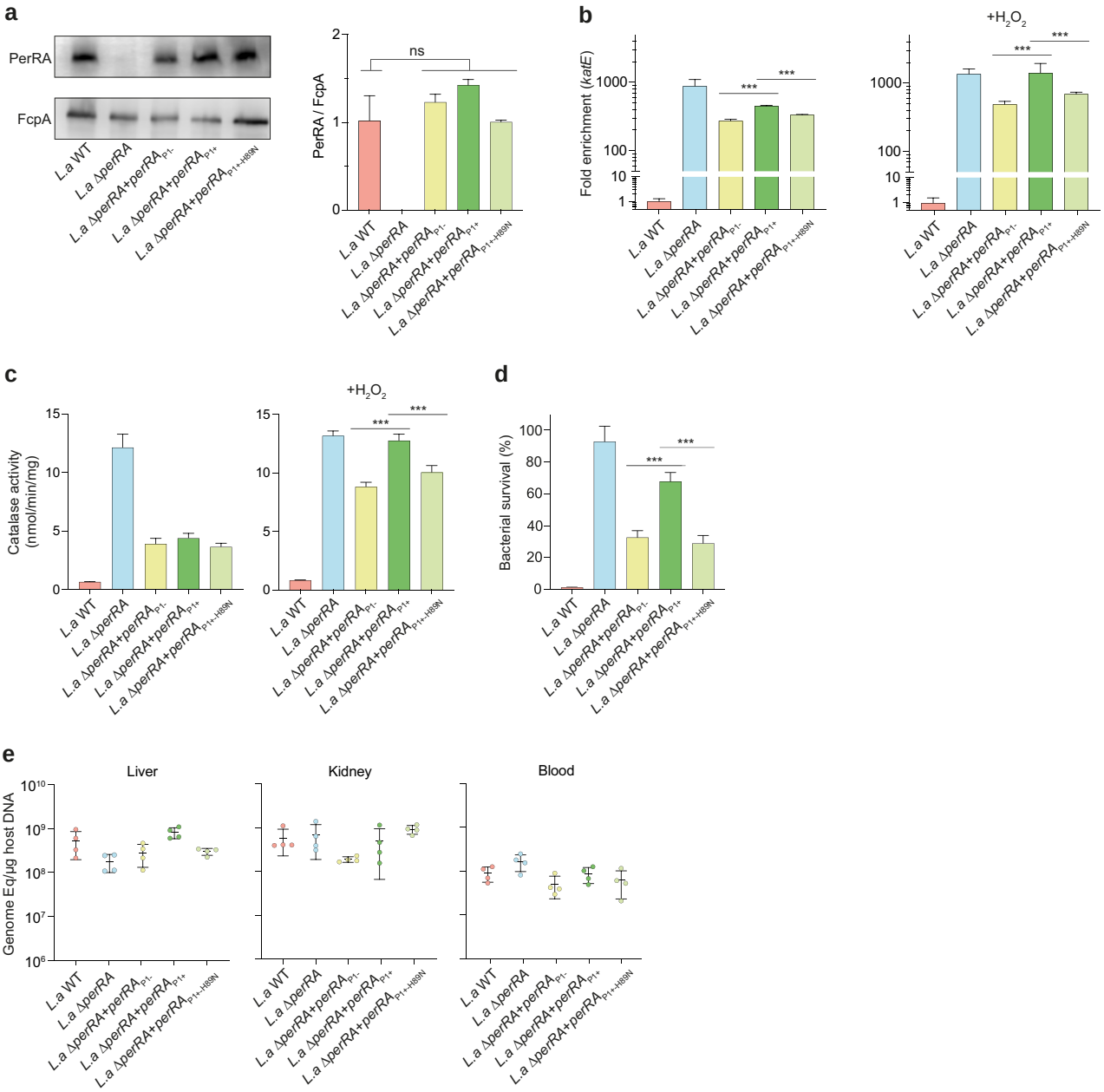

Fig. 8. PerRA from P1+ species improve the tolerance of peroxide stress in P1- species. a) PerRA cellular content of the *L. adleri* (*L.a*) WT and Δ*perRA* mutant complemented with the indicated plasmids was assessed by immunoblot using an anti-FcpA antibody as a control of equal loading (left panel). Densitometric quantification of PerRA was performed and normalized by FcpA amount (right panel) in *L.a* constructs. *L.a* WT and *L.a* Δ*perRA* contain the empty pMaORI-expressing vector; *L.a* Δ*perRA*+*perRA*$_{P1+}$ contains the pMaORI vector bearing the *perRA* ORF of *L.a*; *L.a* Δ*perRA*+*perRA*$_{P1+\text{-}H89N}$ contains the pMaORI vector bearing the *perRA* ORF of *L.a* with the single mutation H89N; *L.a* Δ*perRA*+*perRA*$_{P1-}$ contains the pMaORI vector bearing the *perRA* ORF of *L. adleri*. Data are the means and standard deviation of two independent biological replicates. b) Relative expression of *katE* measured by RT-qPCR in the different *L. adleri* (*L.a*) strains in the absence (left panel) or presence (right panel) of 1 mM of H_2O_2 (30 min exposure). Relative expression levels were normalized to the *flaB2* gene and compared to *L. adleri* WT (*L.a*). c) Measurement of catalase activity in total extracts of the different *L. adleri* (*L.a*) strains in absence (left panel) or in presence (right panel) of 1 mM of H_2O_2 (30 min exposure). d) Survival of the different *L. adleri* (*L.a*) strains upon exposure to 1 mM of H_2O_2 for 30 min. The number of bacteria was enumerated by CFU and compared to H_2O_2-untreated *Leptospira*. e) The virulence of the different *L. adleri* (*L.a*) strains was assessed by infecting hamsters ($n = 4$) by peritoneal route with 10^8 leptospires for each construct. After 1 d of infection, leptospiral load of hamsters infected was assessed by quantitative PCR. For experiments in (a) to (d), one representative experiment of three independent biological replicates is shown. Error bars represent the mean ± SD. Unpaired two-tailed Student's *t* test was used. ***$P < 0.0001$, ns: nonsignificant.

P1- species. Additional divergence in PerRA protein sequence between P1+ and P1- species can be identified, including at positions 88 and 105 (Fig. 6; supplementary fig. S1, Supplementary Material online). PerRAs from P1- species possess charged amino acids at the position 88 and 105 while those of P1+ species possess polar or hydrophobic amino acids. We can infer that such differences might also contribute to the difference in PerRA

activity between P1+ and P1- species. Thorough biochemical characterization of the PerRAs from the two P1 subgroups will be necessary to understand the exact role of these different residues and to test the hypotheses mentioned above.

Catalase has been shown to be required for *Leptospira* virulence (Eshghi et al. 2012). *KatE* gene is present in all species from the P1 subclade but this study suggests that a reduced repression of *katE* resulting in a higher catalase activity may have emerged as an evolutionary advantage for highly virulent species. Together with a higher catalase activity, P1+ species have a reduced expression of a cluster encoding the TonB-dependent energy transduction system. TonB-dependent transporters are often involved in the uptake of iron sources. Accumulation of iron upon peroxide stress worsens oxidative damage because it favors the production of hydroxyl radicals through the Fenton reaction. Lowering iron uptake together with increasing peroxide breakdown by catalase would allow P1+ species to be better equipped to resist the oxidative stress encountered when infecting a host. Using an animal model of acute infection, we were not able to show a clear reduction of virulence of P1+ species or an increase in the ability of P1- species to colonize a mammalian host when *katE* repression was modulated by heterologous complementation. Comparison of the core gene expression across the P1 subclade has identified additional genes that are overexpressed in P1+ species, and some of the factors encoded by these genes might be determinant for *Leptospira* virulence in addition to the catalase. Furthermore, fine-tuning of gene expression control by PerRA may be important at the scale of lineage evolution but too subtle to reproduce in animal models. In addition, it can be implicated in other parts of the life cycle of P1+ species important for their maintenance in the hosts population (chronic colonization of asymptomatic reservoir) or as a way to defend themselves against possible H_2O_2 produced by mammalian commensals competing for the same niche (both not tested here).

In summary, this study demonstrates an accessible bioinformatics pipeline to study intragenus core genes expression change and presents a thorough proof of concept of its utility by studying a regulatory network in pathogenic *Leptospira* species. We demonstrate that we can detect efficiently true transcriptomic changes by showing that a common regulator, PerRA, has different properties between P1+ and P1- species. The reduced PerRA-mediated repression observed in P1+ species facilitates their survival against oxidative stress, probably enabling them to resist harmful oxidants produced by the host's innate immune response and promoting a permissive host infection. This example highlights the importance of combining genomic comparisons with comparative gene expression profiling to fully understand the emergence of complexes phenotypes at the genus scale.

Supplementary Material

Supplementary material is available at *Molecular Biology and Evolution* online.

Acknowledgments

We thank Farah Martin for help with plasmid constructs, Robert Gaultney for RNA extraction.

Funding

This research was supported by the Canadian Institutes of Health Research (CIHR) (#450862) (F.J.V.); by the Institut Pasteur through grant PTR 30-2017 (M.P. and F.J.V.) and by Institut Pasteur & INRS-Centre Armand-Frappier through their Pasteur International Joint Research Units program "LEptospirosis Pasteur NETwork (LePNet)" (M.P. and F.J.V.) and by National Institutes of Health grant P01 AI 168148 (M.P. and F.J.V.). C.N. received a Ph.D. studentship Calmette & Yersin from the Pasteur Network. A.G.G. was funded by a PTR2019-310 grant (NB). S.G.H. was supported by the Pasteur-Paris University (PPU) international PhD program. F.J.V. received a Junior 1 and Junior 2 research scholar salary award from the Fonds de Recherche du Québec Santé. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the article.

Data Availability

The data underlying this article are available in SRA database PRJNA998607, at https://www.ncbi.nlm.nih.gov/bioproject/?term=PRJNA998607. The software Annotator-RNAtor is publicly available on GitHub (https://github.com/BactSymEvol/Annotator-RNAtor) under the GPL-3.0 license.

References

Arnold BJ, Huang IT, Hanage WP. Horizontal gene transfer and adaptive evolution in bacteria. *Nat Rev Microbiol*. 2022:**20**(4):206–218. https://doi.org/10.1038/s41579-021-00650-4.

Bindea G, Mlecnik B, Hackl H, Charoentong P, Tosolini M, Kirilovsky A, Fridman WH, Pages F, Trajanoski Z, Galon J. ClueGO: a Cytoscape plug-in to decipher functionally grouped gene ontology and pathway annotation networks. *Bioinformatics*. 2009:**25**(8):1091–1093. https://doi.org/10.1093/bioinformatics/btp101.

Bryant JM, Brown KP, Burbaud S, Everall I, Belardinelli JM, Rodriguez-Rincon D, Grogono DM, Peterson CM, Verma D, Evans IE, *et al.* Stepwise pathogenic evolution of Mycobacterium abscessus. *Science*. 2021:**372**(6541):eabb8699. https://doi.org/10.1126/science.abb8699.

Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL. BLAST+: architecture and applications. *BMC Bioinformatics*. 2009:**10**(1):421. https://doi.org/10.1186/1471-2105-10-421.

Cantalapiedra CP, Hernandez-Plaza A, Letunic I, Bork P, Huerta-Cepas J. eggNOG-mapper v2: functional annotation, orthology assignments, and domain prediction at the metagenomic scale. *Mol Biol Evol*. 2021:**38**(12):5825–5829. https://doi.org/10.1093/molbev/msab293.

Contreras-Moreira B, Vinuesa P. GET_HOMOLOGUES, a versatile software package for scalable and robust microbial pangenome analysis. *Appl Environ Microbiol*. 2013:**79**(24):7696–7701. https://doi.org/10.1128/AEM.02411-13.

Costa F, Hagan JE, Calcagno J, Kane M, Torgerson P, Martinez-Silveira MS, Stein C, Abela-Ridder B, Ko AI. Global

morbidity and mortality of leptospirosis: a systematic review. *PLoS Negl Trop Dis*. 2015:**9**(9):e0003898. https://doi.org/10.1371/journal.pntd.0003898.

Drew GC, Stevens EJ, King KC. Microbial evolution and transitions along the parasite-mutualist continuum. *Nat Rev Microbiol*. 2021:**19**(10):623–638. https://doi.org/10.1038/s41579-021-00550-7.

Eshghi A, Lourdault K, Murray GL, Bartpho T, Sermswan RW, Picardeau M, Adler B, Snarr B, Zuerner RL, Cameron CE, *et al*. Leptospira interrogans catalase is required for resistance to H2O2 and for virulence. *Infect Immun*. 2012:**80**(11):3892–3899. https://doi.org/10.1128/IAI.00466-12.

Fouts DE, Matthias MA, Adhikarla H, Adler B, Amorim-Santos L, Berg DE, Bulach D, Buschiazzo A, Chang YF, Galloway RL, *et al*. What makes a bacterial species pathogenic?: comparative genomic analysis of the genus Leptospira. *PLoS Negl Trop Dis*. 2016:10(2):e0004403. https://doi.org/10.1371/journal.pntd.0004403.

Hagberg AA, Schult DA, Swart PJ. Exploring network structure, dynamics, and function using networkX. In: Varoquaux G, Vaught T, Millman J, editors. Proceedings of the 7th Python in Science Conference (SciPy2008). Pasadena (CA); Aug 2008. p. 11–15.

Huerta-Cepas J, Forslund K, Coelho LP, Szklarczyk D, Jensen LJ, von Mering C, Bork P. Fast genome-wide functional annotation through orthology assignment by eggNOG-Mapper. *Mol Biol Evol*. 2017:**34**(8):2115–2122. https://doi.org/10.1093/molbev/msx148.

Kebouchi M, Saul F, Taher R, Landier A, Beaudeau B, Dubrac S, Weber P, Haouz A, Picardeau M, Benaroudj N. Structure and function of the Leptospira interrogans peroxide stress regulator (PerR), an atypical PerR devoid of a structural metal-binding site. *J Biol Chem*. 2018:**293**(2):497–509. https://doi.org/10.1074/jbc.M117.804443.

Kristensen DM, Kannan L, Coleman MK, Wolf YI, Sorokin A, Koonin EV, Mushegian A. A low-polynomial algorithm for assembling clusters of orthologous groups from intergenomic symmetric best matches. *Bioinformatics*. 2010:**26**(12):1481–1487. https://doi.org/10.1093/bioinformatics/btq229.

Kristiansson E, Österlund T, Gunnarsson L, Arne G, Joakim Larsson DG, Nerman O. A novel method for cross-species gene expression analysis. *BMC Bioinformatics*. 2013:**14**(1):70. https://doi.org/10.1186/1471-2105-14-70.

Lee J-W, Helmann JD. The PerR transcription factor senses H2O2 by metal-catalysed histidine oxidation. *Nature*. 2006:**440**(7082):363–367. https://doi.org/10.1038/nature04537.

Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*. 2009:**25**(14):1754–1760. https://doi.org/10.1093/bioinformatics/btp324.

Li L, Stoeckert CJ, Roos J, S D. OrthoMCL: identification of ortholog groups for eukaryotic genomes. *Genome Res*. 2003:**13**(9):2178–2189. https://doi.org/10.1101/gr.1224503.

Liao Y, Smyth GK, Shi W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics*. 2014:**30**(7):923–930. https://doi.org/10.1093/bioinformatics/btt656.

Libourel C, Keller J, Brichet L, Cazale AC, Carrere S, Vernie T, Couzigou JM, Callot C, Dufau I, Cauet S, *et al*. Comparative phylotranscriptomics reveals ancestral and derived root nodule symbiosis programmes. *Nat Plants*. 2023:**9**(7):1067–1080. https://doi.org/10.1038/s41477-023-01441-w.

Lo M, Murray GL, Khoo CA, Haake DA, Zuerner RL, Adler B. Transcriptional response of leptospira interrogans to iron limitation and characterization of a PerR homolog. *Infect Immun*. 2010:**78**(11):4850–4859. https://doi.org/10.1128/IAI.00435-10.

Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol*. 2014:**15**(12):550. https://doi.org/10.1186/s13059-014-0550-8.

Maldonado G, Vincent JF, Chenal AT, Veyrier M, J F. CAPRIB: a user-friendly tool to study amino acid changes and selection for the exploration of intra-genus evolution. *BMC Genomics*. 2020:**21**(1):832. https://doi.org/10.1186/s12864-020-07232-3.

Mouville C, Benaroudj N. Survival tests for Leptospira spp. *Methods Mol Biol*. 2020:**2134**:215–228. https://doi.org/10.1007/978-1-0716-0459-5_20.

Nyongesa S, Chenal M, Bernet È, Coudray F, Veyrier FJ. Sequential markerless genetic manipulations of species from the Neisseria genus. *Can J Microbiol*. 2022a:**68**(8):551–560. https://doi.org/10.1139/cjm-2022-0024.

Nyongesa S, Weber PM, Bernet E, Pulido F, Nieves C, Nieckarz M, Delaby M, Viehboeck T, Krause N, Rivera-Millot A, *et al*. Evolution of longitudinal division in multicellular bacteria of the Neisseriaceae family. *Nat Commun*. 2022b:**13**(1):4853. https://doi.org/10.1038/s41467-022-32260-w.

Pappas CJ, Benaroudj N, Picardeau M. A replicative plasmid vector allows efficient complementation of pathogenic Leptospira strains. *Appl Environ Microbiol*. 2015:**81**(9):3176–3181. https://doi.org/10.1128/AEM.00173-15.

Perez JC, Groisman EA. Evolution of transcriptional regulatory circuits in bacteria. *Cell*. 2009:**138**(2):233–244. https://doi.org/10.1016/j.cell.2009.07.002.

Picardeau M. Conjugative transfer between Escherichia coli and Leptospira spp. as a new genetic tool. *Appl Environ Microbiol*. 2008:**74**(1):319–322. https://doi.org/10.1128/AEM.02172-07.

Picardeau M, Brenot A, Saint Girons I. First evidence for gene replacement in Leptospira spp. Inactivation of L. biflexa flaB results in non-motile mutants deficient in endoflagella. *Mol Microbiol*. 2001:**40**(1):189–199. https://doi.org/10.1046/j.1365-2958.2001.02374.x.

Sachs JL, Skophammer RG, Regus JU. Evolutionary transitions in bacterial symbiosis. *Proc Natl Acad Sci U S A*. 2011:**108**(suppl 2):10800–10807. https://doi.org/10.1073/pnas.1100304108.

Said-Salim B, Mostowy S, Kristof AS, Behr MA. Mutations in Mycobacterium tuberculosis Rv0444c, the gene encoding anti-SigK, explain high level expression of MPB70 and MPB83 in Mycobacterium bovis. *Mol Microbiol*. 2006:**62**(5):1251–1263. https://doi.org/10.1111/j.1365-2958.2006.05455.x.

Seemann T. Prokka: rapid prokaryotic genome annotation. *Bioinformatics*. 2014:**30**(14):2068–2069. https://doi.org/10.1093/bioinformatics/btu153.

Sieber M, Traulsen A, Schulenburg H, Douglas AE. On the evolutionary origins of host-microbe associations. *Proc Natl Acad Sci U S A*. 2021:**118**(9):e2016487118. https://doi.org/10.1073/pnas.2016487118.

Tatusov RL, Galperin MY, Natale DA, Koonin EV. The COG database: a tool for genome-scale analysis of protein functions and evolution. *Nucleic Acids Res*. 2000:**28**(1):33–36. https://doi.org/10.1093/nar/28.1.33.

Thibeaux R, Iraola G, Ferres I, Bierque E, Girault D, Soupe-Gilbert ME, Picardeau M, Goarant C. Deciphering the unexplored Leptospira diversity from soils uncovers genomic evolution to virulence. *Microb Genom*. 2018:**4**(1):e000144. https://doi.org/10.1099/mgen.0.000144.

Torgerson PR, Hagan JE, Costa F, Calcagno J, Kane M, Martinez-Silveira MS, Goris MG, Stein C, Ko AI, Abela-Ridder B. Global burden of leptospirosis: estimated in terms of disability adjusted life years. *PLoS Negls Trop Dis*. 2015:**9**(10):e0004122. https://doi.org/10.1371/journal.pntd.0004122.

Vincent AT, Schiettekatte O, Goarant C, Neela VK, Bernet E, Thibeaux R, Ismail N, Mohd Khalid MKN, Amran F, Masuzawa T, *et al*. Revisiting the taxonomy and evolution of pathogenicity of the genus Leptospira through the prism of genomics. *PLoS Negl Trop Dis*. 2019:**13**(5):e0007270. https://doi.org/10.1371/journal.pntd.0007270.

Waterhouse AM, Procter JB, Martin DMA, Clamp M, Barton GJ. Jalview version 2—a multiple sequence alignment editor and

analysis workbench. *Bioinformatics*. 2009:**25**(9):1189–1191. https://doi.org/10.1093/bioinformatics/btp033.

Zavala-Alvarado C, Benaroudj N. The single-step method of RNA purification applied to Leptospira. *Methods Mol Biol*. 2020:**2134**:41–51. https://doi.org/10.1007/978-1-0716-0459-5_5.

Zavala-Alvarado C, Sismeiro O, Legendre R, Varet H, Bussotti G, Bayram J, GH S, Rey G, Coppee JY, Picardeau M, *et al.* The transcriptional response of pathogenic Leptospira to peroxide reveals new defenses against infection-related oxidative stress. *PLoS Pathog.* 2020:**16**(10):e1008904. doi:10.1371/journal.ppat.1008904.