
REGULATION OF THE *Acanthamoeba castellanii* GENOME UPON INFECTION BY *Legionella pneumophila*

A PREPRINT

August 24, 2021

Cyril Matthey-Doret^{¶,1,2}, Morgan Colp^{¶,3}, Pedro Escoll Guerrero⁴, Bruce Curtis⁵, Matt Sarrasin⁵, B. Franz Lang⁵, Michael W. Gray⁶, John M. Archibald³, Carmen Buchrieser³, Romain Koszul^{1,*}

1 Institut Pasteur, Spatial Regulation of Genomes unit, CNRS, UMR 3525, C3BI USR 3756, Paris, France

2 Sorbonne Université, Collège Doctoral, F-75005 Paris, France

3 Department of Biochemistry and Molecular Biology, Dalhousie University, Halifax, Nova Scotia, Canada

4 Institut Pasteur, Biology of Intracellular Bacteria Unit, CNRS, UMR 3525, Paris, France

5 Robert Cedergren Centre for Bioinformatics and Genomics, Département de Biochimie, Université de Montréal, Montreal, QC, Canada

6 Department of Biochemistry and Molecular Biology and Centre for Comparative Genomics and Evolutionary Bioinformatics, Sir Charles Tupper Medical Building, Dalhousie University, 5850 College Street, Halifax, Nova Scotia, B3H 4R2, Canada

¶ These authors contributed equally.

ABSTRACT

The unicellular amoeba *Acanthamoeba castellanii* is ubiquitous in aquatic environments. This protozoa may host bacterial endosymbionts and intracellular parasites. Some of these are human pathogens, such as Chlamydia or Legionella. The current reference genome of *A. castellanii* is based on strain Neff, by far the most extensively studied strain of this species, and is resolved at scaffold-level. However, little is known about *A. castellanii* genome variability among strains or the influence infection by pathogens may exert on the regulation of the protozoan host genome. Here we generated unique chromosome-level assemblies of *A. castellanii* str. Neff and *A. castellanii* str. C3 using a hybrid approach combining long reads, Hi-C and shotgun sequencing. Investigating their genomes revealed an unexpected high conservation among the two strains as only XX different. We also describe the spatial organization of the *A. castellanii* genome and how it changes during infection by the bacterium *L. pneumophila*.

Keywords Genomics · Assembly · Comparative genomics · Hi-C · Host-parasite

1 Introduction

Acanthamoeba species, such as *Acanthamoeba castellanii* are aerobic, unicellular, free living amoeba, present throughout the world in soil and nearly all aquatic environments [1]. *Acanthamoeba* species are also opportunistic pathogens that can cause serious infections in immunocompromised humans. In their natural environment, they are predators feeding on bacteria, that may also phagocytize yeasts or algae. Over time, *Acanthamoeba* also became reservoirs of microorganisms and viruses, including human pathogens, which have adapted to survive inside these cells and resist digestion, persist or even replicate as intracellular parasites. At least 15 different bacterial species, two archaea and

several eukaryotes and viruses have been shown to interact with *Acanthamoeba* in the environment and may even co-exist at the same time in the same host cell [2].

Acanthamoeba and some bacteria thus share a common evolutionary history, and the question of whether this history is reflected in their metabolism, and therefore in their genome, has arisen. The genome of *A. castellanii* is rich in lateral gene transfers, encodes a large variety of enzymes to degrade various substrates (cellulose, biofilms, ...) and contains transposable elements [3]. Another unique feature of this genome is that the 5S rDNA genes are dispersed through the genomic sequence, whereas other eukaryotic organisms normally have a cluster of tandem 5S rDNA genes [4].

Although it was observed early on that bacteria could resist digestion of free-living amoebae [5], it was not until the discovery that *Legionella pneumophila* replicated in amoebae that the bacteria-amoeba relationship began to be studied in depth [6]. *L. pneumophila* is the agent responsible for the legionnaire's disease, a severe pneumonia that can be fatal if not treated promptly. In addition, many species of amoeba species have the ability to form highly resistant cysts in hostile environments, providing shelter for their intracellular parasites [7]. For instance, *L. pneumophila* is suspected to survive water disinfection treatments, and contaminating water distribution systems, thanks to cysts protection [8, 9, 10]. From these contaminated water sources, *L. pneumophila* can reach and replicate within human lungs via aerosols.

Furthermore, *L. pneumophila* has the ability to escape the lysosomal degradation pathway of both *A. castellanii* and human alveolar macrophages through the formation of a protective vacuole (the Legionella-containing vacuole or LCV) where it multiplies to high numbers. To establish the LCV and replicate, *L. pneumophila* secretes over 300 effector proteins via a type four secretion system (T4SS) called Dot/Icm [11] into the host cytoplasm to manipulate the host pathways and redirect nutrients to the LCV [12, 13]. In the early stages of infection, many of these proteins target the host secretory pathway, including several small GTPases, to recruit the endoplasmic reticulum [14]. During the intracellular cycle, a wider range of processes, including membrane trafficking, cytoskeleton dynamics or signal transduction pathways, are targeted by these effectors [15, 16]. *L. pneumophila* will also directly alter the genome of its host, by modifying epigenetic marks of the host genome in human macrophages and in *A. castellanii*, by secreting an effector named RomA that encodes a histone methyltransferase secreted in the host cell and targeted to the host nucleus. RomA trimethylates K14 of histone H3 [17] genome wide, leading to transcriptional changes that modulate the host response in favor of the survival of the bacteria [17]. Concomitantly, *L. pneumophila* infection leads to genome-wide changes in gene expression [18]. In many eukaryotes, gene regulation is intertwined with the tridimensional organization of chromosomes, and whether chromatin loops, self-interacting domains, and/or active/inactive compartments play roles in cellular development, response to environmental changes, or processes such as DNA repair, are actively being studied [19, 20]. Therefore, the infection of *A. castellanii* by *L. pneumophila* provides an amenable model to investigate how an intracellular bacterial infection may affect the regulation of chromosome folding, and its consequences, in a eukaryotic host.

The investigation of the genome organization and regulatory states of *A. castellanii* in response to infection requires a well assembled genome. The current genome reference of *A. castellanii* (NEFF-v1, [3] is based on the Neff strain, isolated from soil in California in 1957 [21], and by far the most widely used in labs. This assembly is fragmented into 384 scaffolds, which makes chromosome-level analyses difficult, if not impossible. In addition, some general features of the *A. castellanii* genome, such as the number of chromosomes or the ploidy, remain undetermined. In addition, many teams investigating bacteria-amoeba interaction favor the "C3" strain (ATCC 50739), isolated from a drinking water reservoir in Europe in 1994 and identified as mice pathogen [22], for its higher sensitivity to infection. This, despite scarce genomic information and little information regarding its phylogenetic distance to the Neff strain. Notably, the extent of genome conservation between these two *A. castellanii* strains that have been cultivated for several decades and were isolated from different ecological niches is unknown. The lack of genomic resources hampers the investigation of the factors determining the pathogen susceptibility of different *A. castellanii* strains, as well as the application of genomewide omic approaches.

In this work, we aimed at characterising the response to infection of *A. castellanii* C3 strain by *L. pneumophila* through the prism of the three-dimensional organisation of the amoeba genome. This analysis required the generation of a high quality reference genome sequence of the C3 strain, which we then compared to a new assembly of the Neff reference strain also presented in this paper. Illumina shotgun, Nanopore long read, and Hi-C were used to generate near chromosome-level assemblies of both strains. Our new Neff and C3 assemblies have a (gap-excluded) sequence divergence of 6.9%. We find that core metabolism genes are largely conserved between the two strains and differences in gene content are mostly associated with signal transduction. Using the C3 assembly, RNA-seq and Hi-C, we were then able to analyze the genome folding and expression changes of *A. castellanii* in response to the infection by *L. pneumophila*. We found infection-dependent chromatin loops to be enriched in signal transduction and phosphorylation.

strain	step	length	scaffolds	N50	L50	N90	L90
C3	01-flye	46527779	368	406024	28	82982	122
C3	02-hypo	46287494	368	402530	28	82062	122
C3	03-instagraal	46287494	253	1424952	13	895233	29
C3	04-pilon	46273248	225	1424952	13	895233	29
C3	05-manual	46256395	227	1400265	14	895233	31
C3	06-rcorr	46255455	227	1400134	14	895207	31
Neff	01-flye	44412593	237	1159072	15	141936	51
Neff	02-hypo	44503078	237	1161784	15	142434	51
Neff	03-instagraal	44503078	179	1460606	13	965190	28
Neff	04-pilon	44504078	179	1460763	13	965290	28
Neff	05-manual	43957095	178	1332098	14	911389	29
Neff	06-rcorr	43956874	178	1332098	14	911382	29

Table 1: Assembly statistics for *A. castellanii* at each pipeline step. For both strains, general assembly metrics are shown alongside the name of the tool used in that step and its position within the pipeline. Manual denotes manually curated changes.

2 Results

2.1 Genome assembly

We used a combination of ONT long reads, Hi-C and shotgun sequencing to reassemble the genome at chromosome level, with 90% of the genome comprised in 28 scaffolds, in contrast to the number of chromosomes previously estimated to be around 20 from pulsed-field gel electrophoresis [23]. For both the Neff and C3 strains, we first generated a raw *de-novo* assembly using Oxford Nanopore long reads, of 44.4 and 46.5 Mb in size, respectively (Table 1). To account for the error prone nature of long reads, we combined them with paired-end shotgun Illumina sequences to polish this first draft assembly using HyPo [24]. The polished assembly was then scaffolded with long range Hi-C contacts using our probabilistic program instaGRAAL, which exploits a Markov Chain Monte Carlo algorithm to swap DNA segments until the likeliest scaffold is achieved [25]. Following the post-scaffolding polishing step of the program (see [25]), the final assemblies of both strains displayed a better contiguity, completion and mapping statistics than the previous version of the assembly, with the cumulative length of the scaffolds quickly reaching a plateau (Fig. 1a). The assemblies of both strains are also slightly longer, with a smaller number of contigs than the previous Neff assembly (NEFF-v1) (Fig. 1b). The BUSCO-completeness of the assembly is also improved, with 90.6% (Neff) and 91.8% (C3) complete eukaryotic universal single copy orthologs, compared to 77.6% for NEFF-v1. Hi-C contact maps consist in a convenient readout to explore large misassemblies in genome sequences [26]. While this allowed us to manually address major unambiguous misassemblies, a number of visible misassemblies remain in complex regions such as repeated sequences near telomeres or rDNAs. In C3, there are also a few (at least 5) interchromosomal misassemblies which appear to be heterozygous and cannot be resolved without a phased genome. We also found shotgun coverage to be highly heterogeneous between scaffolds, which is suggestive of aneuploidies (Fig. S1).

2.2 Comparison of Neff and C3 strains

Generating the genome assemblies of the two strains offered us the opportunity to compare them. We used Orthofinder to compare the gene content of the C3 and Neff strains. We found xxx (mannose-binding lectin + gene content details) xx.

2.3 Spatial organisation of the *A. castellanii* genome

To our knowledge, no Hi-C contact maps have been generated from species of the Amoebozoa clade. Therefore, the Hi-C reads used to generate the chromosome scale scaffolding of two *A. castellanii* strains also offer the opportunity to expose the average genome folding in a species of this clade. Hi-C reads were realigned along the new assemblies of both the C3 and Neff strains to generate genomewide contact maps. Visualising the Hi-C contact maps of both strains reveals that *A. castellanii* chromosomes are well resolved in our assembly (Fig. 2). In Neff, the highest intensity contacts are concentrated on the main diagonal, suggesting an absence of large-scale misassemblies. On the other hand, the C3 assembly retains a few misassembled blocks, mostly in the rDNA region where tandem repeats could not be resolved correctly. However, for both strains the genomewide contact maps reveal a grid-like pattern, with contact enrichment between chromosome extremities resulting in discrete dots. These contacts can be interpreted as a clustering

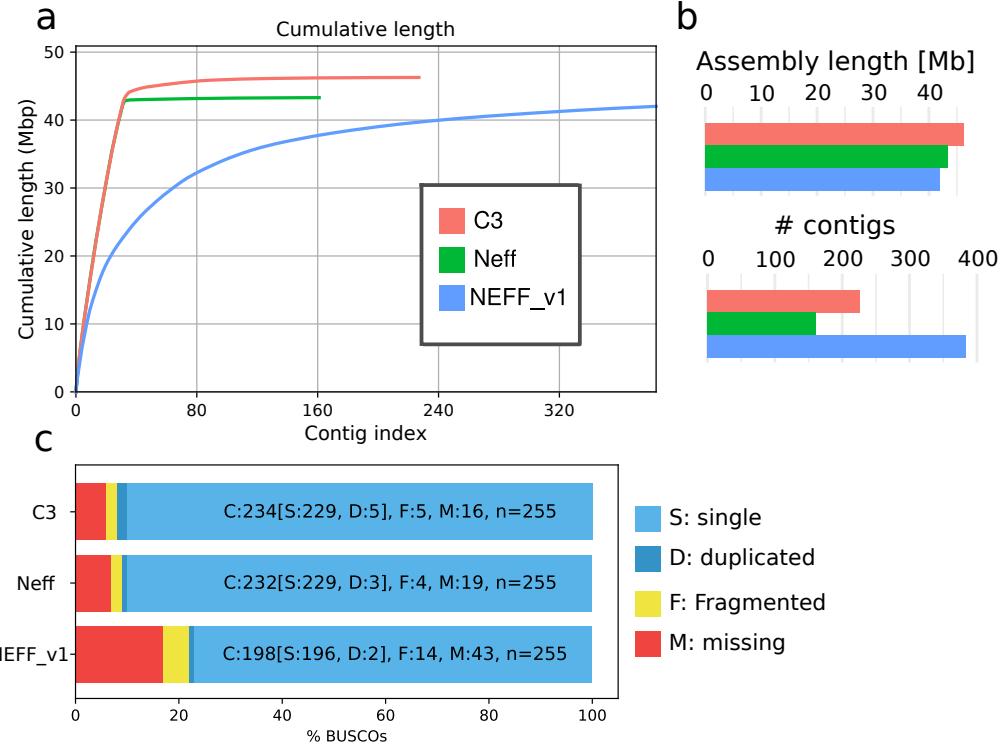


Figure 1: Assembly statistics for *A. castellanii*. Comparison genome assemblies for strains C3 and Neff, versus the previous NEFF-v1 genome assembly [3]. **a**, Cumulative length plot showing the relationship between number of contigs (largest to smallest) and lenght of the assembly. **b**, General continuity metrics. **c**, BUSCO statistics showing the status of universal single copy orthologs in eukaryotes for each assembly.

of the telomeres, or subtelomeres, of the different chromosomes (Fig. 2). Additionally we found ~ 100 copies of 5S ribosomal DNA genes dispersed on most chromosomes for both strains, and 18S/28S genes show increased contacts with subtelomeres (Fig. 2).

In addition to large, trans- subtelomeric contacts, we also sought for intrachromosomal chromatin 3D structures in the contact maps using Chromosight, a pattern-detection program [27]. For both strains, the program pointed at arrays of chromatin loops along chromosomes, as well as boundaries separating chromatin domains (Fig. 2b). Most chromatin loops are regularly spaced, with a typical size of ~ 20 kb (Fig. 2c). The chromatin domains correspond to discrete squares along the diagonal (Fig. S3a). We overlapped all genes in the C3 strain with the domain borders we detected from Hi-C data and measured their base expression using RNAseq we generated from that strain (methods). We found that the genes overlapping domain boundaries are overall more highly expressed compared to the genome average (Fig. S2c). In addition, the analysis showed that gene expression is negatively correlated with the distance to the closest domain border (Fig. S2d). We performed the same comparison using chromatin loop anchors instead of domain borders. To a lesser extent, genes overlapping chromatin loops are also associated with a higher expression (Fig. S2a), although it is not correlated with the distance from the closest loop (Fig. S2b). Altogether, these results suggest that the chromatin structures observed in *cis* are both associated with gene expression, however the association between gene expression and chromatin loops basis is likely due to their co-localization with domain borders (Fig. S2e). Altogether, these results are reminiscent of the organization of microorganisms chromosomes into small micro domains corresponding to expressed genes [28, 29].

2.4 Effect of *L. pneumophila* infection on the genome of *A. castellanii*

Having generated near complete assemblies allowed us to tackle the investigation of the effect of *L. pneumophila* infection on the 3D folding and transcription of *A. castellanii* C3 strain genome. To do so, we harvested cultured *A. castellanii* cells before and 5 hours following infection by *L. pneumophila* strain Paris (Methods). The cells were processed using Hi-C and RNA-seq (Methods), and the resulting reads aligned against the reference genome to assess for changes in the genome structure and the host transcription program, respectively. The RNAseq was performed

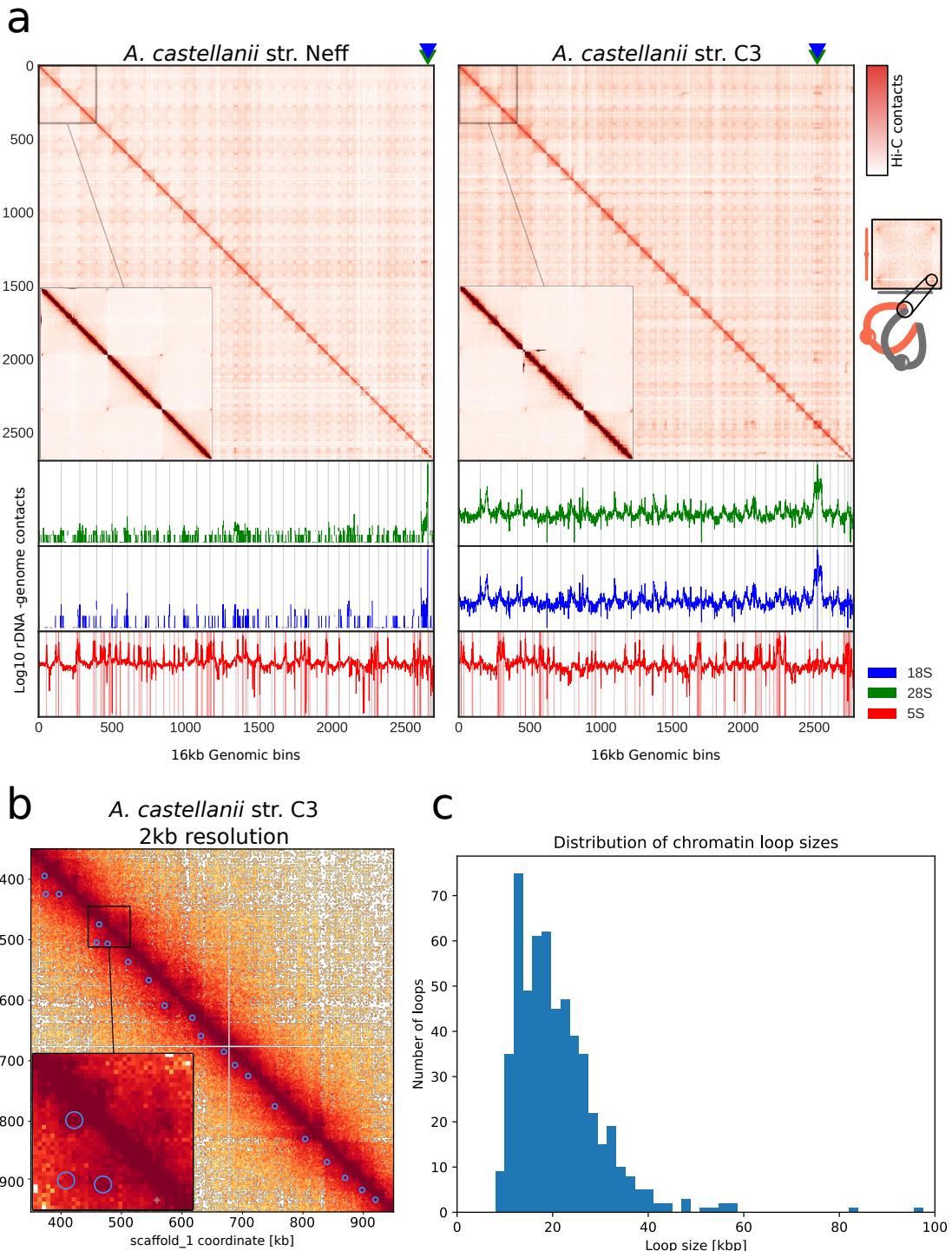


Figure 2: Spatial organisation of the *A. castellanii* genome. **a**, Top: Whole genome Hi-C contact maps of the Neff (left) and C3 (right) strains, with a magnification of the 3 largest scaffolds. The genome is divided into 16 kb bins, and each pixel represents the contact intensity between a pair of bins. Each scaffold is visible as a red square on the diagonal. In both strains, there is an enrichment of inter-scaffold contacts towards telomeres, suggesting a spatial clustering of telomeres, as shown on the model in the right margin. Bottom: 4C-like representation of spatial contacts between rDNA and the rest of the genome. Scaffolds are delimited by grey vertical lines. Contacts of all rDNAs are enriched towards telomere. The genomic position of 18S and 28S sequences are highlighted with triangles on the top panel and the occurrences of 8S rDNA sequences are shown with vertical red lines on the bottom panel. **b**, High resolution the contact map for a region of the C3 genome showing chromatin loops detected by Chromosight as blue circles. **c**, Size distribution of chromatin loops detected in the C3 strain.

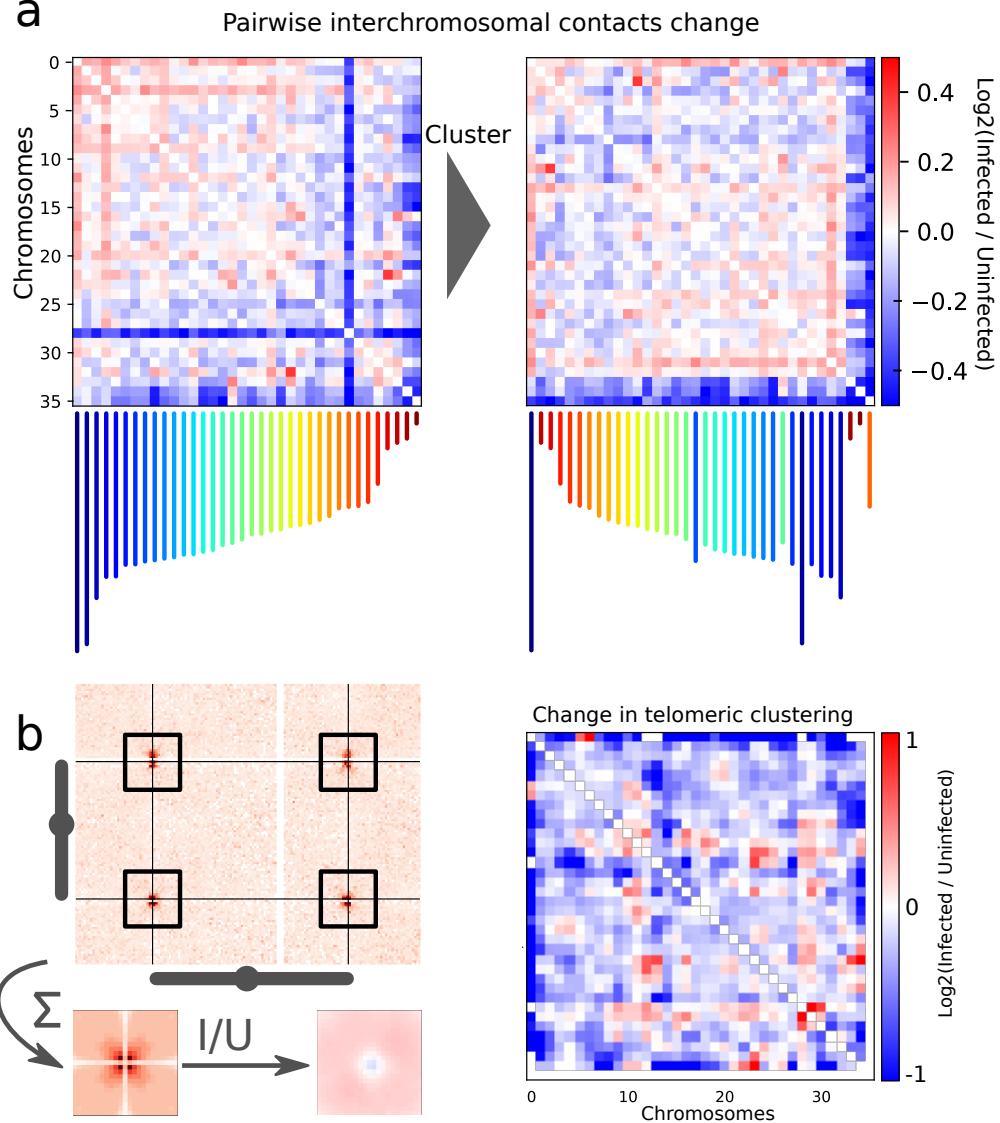


Figure 3: Changes in trans-chromosomal contacts between *A. castellanii* chromosomes following *L. pneumophila* infection. **a**, Average contact change during infection between each pair of chromosomes before (left) and after (right) clustering chromosomes by interchromosomal contact profile. **b**, Inter-telomeric contacts between pairs of chromosomes are summed to generate a pileup, which can also be visualized as a ratio between two conditions (left). Change in telomeric pattern intensity during infection for all pairs of chromosomes (right). The intensity is the Pearson correlation coefficient between the telomeric pileup and each telomeric window. Each value in the matrix represents a pair of chromosomes, consisting of the average of 4 telomeric windows. Intrachromosomal windows are excluded.

in triplicates, and the Hi-C in duplicates (Methods). To measure changes in trans-chromosomal contacts, we merged replicates contact maps and applied the serpentine adaptative binning to improve the signal-to-noise ratio [30]. We then computed average interactions between each pair of chromosomes before and after infection. For each pair of chromosome, we then use the log ratio of infected over uninfected average contacts. Following infection a global decrease in trans-subtelomeric contacts was observed, suggesting a slight de-clustering of chromosome ends (Fig. 3b). In addition, the scaffold bearing 18S and 28S rDNA (scaffold_29) genes, as well as two other small scaffolds (35 and 36) displayed weaker interactions with other scaffolds during infection (Fig. 3a).

We then assessed whether the behavior of cis-contacts changes during infection. First, we computed the average contact frequencies according to genomic distance $p(s)$ (Methods), which is a convenient way to unveil variations in the

compaction state of chromatin [31]. The p(s) curves show a global increase in long range contacts following infection (Fig. S4b). The strengths of chromatin loops and domain borders were quantified using Chromosight [27], before and 5h after infection. However, no significant average increase or decrease in the intensity of these structures (Fig. S4a) was identified when computed over the whole genome. To focus on infection-dependent chromatin structures, we filtered the top 20% detected patterns showing the strongest change in Chromosight score during infection (either appearing or disappearing). We performed a GO term enrichment analysis for genes associated with infection-dependent chromatin loops (methods). A significant enrichment for Rho GTPase and phosphorelay signal transduction, protein catabolism and GPI biosynthesis was found (Fig. S6a). Strongest loop changes were associated with genes including Rho GTPase, GOLD and SET domains as well as genes containing leucine-rich repeats and ankyrin repeats (Fig. S7).

We followed the same procedure from domain borders and found that genes associated with infection-dependent domain borders were significantly enriched in amino acid transport, cyclic nucleotide biosynthetic process, protein modification and deubiquitination (Fig. S6a). The presence of several other enriched metabolism-related terms is consistent with the fact that borders are generally associated with highly transcribed genes [32].

Using our infection RNAseq we also found most the expression of genes to be deregulated at 5h post infection compared to uninfected cells (Fig. S5a). To investigate the relationship between this deregulation and chromatin structures, we assigned the closest domain border to each gene and compared their expression and border score changes during infection. For the majority of genes, we found border intensity to be decorrelated from gene expression changes (Fig. S5b). Only genes undergoing extreme expression changes during infection reflected changes in associated borders (Fig. S5c). This suggests that insulation domains in *A. castellanii* have a lesser importance in dictating gene expression programs than in mammals.

Recently, Li et al. [18] investigated gene expression changes at 3, 8, 16 and 24h after infection of *A. castellanii* Neff by *Legionella pneumophila*. They also reported a global deregulation at 8h compared to cells infected with $\Delta dotA$ mutant bacteria. To further validate our finding that chromatin domains are not units of regulation in *A. castellanii*, we used their expression results and lifted over the gene annotations to our C3 assembly using liftoff [33]. This allowed us to compute the co-expression between gene pairs during infection (i.e. expression correlation). We found that genes pairs sharing the same chromatin domain did not have a higher co-expression than gene pairs from different domains at similar genomic distances (Fig. S3).

3 Discussion

The heterogeneous coverage we found across scaffolds would be consistent with previous findings that *A. castellanii* has a high but variable ploidy of approximately 25n. [34]. Considering the biology of *A. castellanii*, with suspected amitosis [35] and probable aneuploidies, a dispersion of 5S ribosomal DNA across all chromosomes would make sense, to ensure a consistent copy number of 5S rDNA in daughter cells.

It was previously estimated that *A. castellanii* has 24 copies of the rDNA gene per haploid genome [36], but we found 4 times as many in both strains. The decrease in interchromosomal contacts with rDNA-containing scaffold could reflect to an alteration in the nucleolus structure. This could be caused by a global increase in translational activity during infection, which would be consistent with the global transcription shift observed in RNAseq.

At first glance, the contact maps show a clustering of subtelomeric regions, but do not display a Rabl conformation, where centromeres cluster to the spindle-pole body [37]. However, the precise positions of centromeres will be needed to verify they don't co-localize with subtelomeric regions. It was previously shown that *L. pneumophila* infection halts the host cell division, and is associated with a decrease of mRNA of the *A. castellanii* CDC2b gene [38]. The large scale 3D changes we observe in chromatin compaction (Fig. S4b) and interchromosomal contacts (Fig. 3) are reminiscent of cell cycle changes in yeast and could suggest that the bacterium stops its host's cell cycle at a specific checkpoint.

We identified array of regularly spaced chromatin loop of approximately 20kb in size. This also happens to be the range of the chromatin loops observed in *S. cerevisiae* during the G2/M stage [39]. This similarity in terms of regularity and size suggests that chromatin loops in *A. castellanii* may serve a similar purpose of chromosome compaction for cell division as in yeast. Our finding that DNA loops basis overlap highly expressed genes is also concordant with observations made in yeast or other species [40], and result presumably from their role in blocking processing SMC complexes [41].

Unlike previously shown in Drosophila [42], we found no increase in co-expression for genes sharing the same contact domain in *A. castellanii*. This suggests chromatin domains may be caused by highly transcribed genes, and are do not act as units of regulation as in other multicellular eukaryotes.

4 Methods

4.1 Strains and growth conditions

4.2 Infection timecourse

A. castellanii str. Neff and C3 were grown on ... at ... they were infected by *L. pneumophila* str. Philadelphia with MOI 10 after 5h. DNA was then extracted and crosslinked.

4.3 Library preparations

4.3.1 Hi-C

Hi-C libraries were prepared according to the Arima protocol using only the DpnII enzyme. Libraries were sequenced at 2x35bp on an Illumina NextSeq machine.

4.3.2 Shotgun

Shotgun libraries were sequenced by Novogene at 2x150bp on an Illumina Novaseq machine.

4.3.3 RNA-seq

RNAseq libraries were prepared using the stranded mRNA Truseq kit from Illumina and sequenced in single-end mode at 1x150bp on an Illumina NextSeq machine.

4.3.4 Nanopore sequencing

A. castellanii from str. Neff and C3 was extracted according to the Nanopore protocol for high molecular weight gDNA extraction from cell lines. Nanopore libraries were prepared with the ligation sequencing kit LSKQ109, flowcell model MIN106D R9. Basecalling was performed using Guppy v2.3.1-1.

4.4 Genome assembly

Nanopore reads were filtered using filtlong v0.2.0 with default parameters to keep the best 80% reads according to length and quality. Illumina shotgun libraries were used as reference for the filtering. A *de-novo* assembly was generated from the raw (filtered) Nanopore long reads using flye v2.3.6 with 3 iterations. The resulting assembly was polished using both Nanopore and Illumina reads with HyPo v1.0.1.

Contigs from the polished assembly bearing more than 60% of their sequence or 51% identity to the mitochondrial sequence from the NEFF_v1 assembly were separated from the rest of the assembly to prevent inclusion of mitochondrial contigs into the nuclear genome during scaffolding.

Polished nuclear contigs were scaffolded with Hi-C reads using instagraal v0.1.2 with default parameters. Instagraal-polish was then used to fix potential errors introduced by the scaffolding procedure. Mitotic contigs were then added at the end of the scaffolded assembly and the final assembly was polished with the Illumina shotgun library using 2 rounds of pilon polishing. The resulting assembly was edited manually to remove spurious insertion of mitochondrial contigs in the scaffold and other contaminants. The final assembly was polished again using pilon with Rcorrector-corrected reads [43].

Minimap2 v2.17 was used for all long reads alignments, and bowtie2 v2.3.4.1 for short reads alignments.

4.5 Genome annotation

The Neff and C3 assemblies were annotated ...

Functional annotations were added using funannotate v1.5.3. Repeated sequences were masked using repeatmasker. Predicted proteins were fed to Interproscan v5.22, Phobius v1.7.1 and EggNOG-mapper v2.0.0 were used to generate functional annotations. Ribosomal DNA genes were annotated separately using RNAmmer v1.2 with HMMER 2.3.2.

The funannotate-based script "func_annot_from_gene_models.sh" used to add functional annotations to existing gene models is provided in the zenodo record xx and on the associated github repository as described in the code availability section.

4.6 Comparative analyses

Proteomes of 11 other amoeba species and 3 intracellular bacteria associated to *A. castellanii* were retrieved from NCBI. Redundant sequences in each proteomes with more than 95% identity were filtered out using CDHit v4.8.1. The protein sequences were fed to orthofinder v2.3.3 to obtain orthogroups of proteins, and a matrix of presence-absence was built to identify potential HGT events present in *A. castellanii* and intracellular bacteria, but not other amoeba species.

The pipeline for comparative genomics analyses is available on Github at: ...

To compute proportion of substituted positions in aligned segments between C3 and Neff strains, the two genomes were aligned using minimap2 with the asm20 preset and -c flag. The gap-excluded sequence divergence (mismatches / (matches + mismatches) was then computed in each primary alignment and the average of divergences (weighted by segment lengths) was computed.

4.7 Hi-C analyses

Reads were aligned with bowtie2 v2.4.1 and Hi-C matrices were generated using hicstuff v3.0.1 (<https://www.github.com/koszullab/hicstuff>). For all comparative analyses, matrices were downsampled to the same number of contacts using cooltools (<https://www.github.com/mirnylab/cooltools>) and balancing normalization was performed using the ICE algorithm [44]. Loops and domain borders were detected using Chromosight v1.6.1 [27] using the merged replicates at a resolution of 2kbp. We measured the intensity changes during infection using pareidolia v0.6.1 (<https://www.github.com/koszullab/pareidolia>) on 3 pseudo replicates generated by sampling the merged contact maps, as described in [45]. This was done to account for contact coverage heterogeneity across replicates.

5 Code availability

The analyses are packaged into the following snakemake pipelines available on github.

- Hybrid genome assembly: https://github.com/cmdoret/Acastellanii_hybrid_assembly
- Functional annotation of *A. castellanii*: https://github.com/cmdoret/Acastellanii_genome_annotation
- Analyses of genomic features in *A. castellanii*: https://github.com/cmdoret/Acastellanii_genome_analysis
- Changes during infection by Legionella: https://github.com/cmdoret/Acastellanii_legionella_infection.

All aforementioned repositories have been linked to Zenodo record xx available at: xx

References

- [1] Salvador Rodríguez-Zaragoza. Ecology of Free-Living Amoebae. *Critical Reviews in Microbiology*, 20(3):225–241, January 1994. Publisher: Taylor & Francis _eprint: <https://doi.org/10.3109/10408419409114556>.
- [2] Ascel Samba-Louaka, Vincent Delafont, Marie-Hélène Rodier, Estelle Cateau, and Yann Héchard. Free-living amoebae and squatters in the wild: ecological and molecular features. *FEMS Microbiology Reviews*, 43(4):415–434, July 2019.
- [3] Michael Clarke, Amanda J Lohan, Bernard Liu, Ilias Lagkouvardos, Scott Roy, Nikhat Zafar, Claire Bertelli, Christina Schilde, Arash Kianianmomeni, Thomas R Bürglin, Christian Frech, Bernard Turcotte, Klaus O Kopec, John M Synnott, Caleb Choo, Ivan Paponov, Aliza Finkler, Chris Heng Tan, Andrew P Hutchins, Thomas Weinmeier, Thomas Rattei, Jeffery SC Chu, Gregory Gimenez, Manuel Irimia, Daniel J Rigden, David A Fitzpatrick, Jacob Lorenzo-Morales, Alex Bateman, Cheng-Hsun Chiu, Petrus Tang, Peter Hegemann, Hillel Fromm, Didier Raoult, Gilbert Greub, Diego Miranda-Saavedra, Nansheng Chen, Piers Nash, Michael L Ginger, Matthias Horn, Pauline Schaap, Lis Caler, and Brendan J Loftus. Genome of Acanthamoeba castellanii highlights extensive lateral gene transfer and early evolution of tyrosine kinase signaling. *Genome Biology*, 14(2):R11, 2013.
- [4] Michael G. Zwick, Melanie Wiggs, and Marvin R. Paule. Sequence and organization of 5S RNA genes from the eukaryotic protist Acanthamoeba castellanii. *Gene*, 101(1):153–157, May 1991.
- [5] W. Drozanski. Fatal bacterial infection in soil amoebae. *Acta Microbiologica Polonica* (1952), 5(3-4):315–317, 1956.

- [6] T J Rowbotham. Preliminary report on the pathogenicity of *Legionella pneumophila* for freshwater and soil amoebae. *Journal of Clinical Pathology*, 33(12):1179–1183, December 1980.
- [7] S. Kilvington and Jackie Price. Survival of *Legionella pneumophila* within cysts of *Acanthamoeba polyphaga* following chlorine exposure. *Journal of Applied Bacteriology*, 68(5):519–525, May 1990.
- [8] Isabelle Pagnier, Didier Raoult, and Bernard La Scola. Isolation and identification of amoeba-resisting bacteria from water in human environment by using an *Acanthamoeba polyphaga* co-culture procedure. *Environmental Microbiology*, 10(5):1135–1144, May 2008.
- [9] Agnes Lasheras, Helene Boulestreau, Anne-Marie Rogues, Celine Ohayon-Courtes, Jean-Claude Labadie, and Jean-Pierre Gachie. Influence of amoebae and physical and chemical characteristics of water on presence and proliferation of *Legionella* species in hospital water systems. *American Journal of Infection Control*, 34(8):520–525, October 2006.
- [10] Masanari Ikeda and Eiko Yabuuchi. Ecological Studies of *Legionella* Species: I. Viable Counts of *Legionella pneumophila* in Cooling Tower Water. *Microbiology and Immunology*, 30(5):413–423, May 1986.
- [11] Tomoko Kubori and Hiroki Nagai. The Type IVB secretion system: an enigmatic chimera. *Current Opinion in Microbiology*, 29:22–29, February 2016.
- [12] Ralph R. Isberg, Tamara J. O'Connor, and Matthew Heidtman. The *Legionella pneumophila* replication vacuole: making a cosy niche inside host cells. *Nature Reviews Microbiology*, 7(1):13–24, January 2009.
- [13] Alexander W. Ensminger. *Legionella pneumophila*, armed to the hilt: justifying the largest arsenal of effectors in the bacterial world. *Current Opinion in Microbiology*, 29:74–80, February 2016.
- [14] Anna Leoni Swart, Laura Gomez-Valero, Carmen Buchrieser, and Hubert Hilbi. Evolution and function of bacterial RCC1 repeat effectors. *Cellular Microbiology*, 22(10):e13246, October 2020.
- [15] Andree Hubber and Craig R. Roy. Modulation of host cell function by *Legionella pneumophila* type IV effectors. *Annual Review of Cell and Developmental Biology*, 26:261–283, 2010.
- [16] Jiazhang Qiu and Zhao-Qing Luo. *Legionella* and *Coxiella* effectors: strength in diversity and activity. *Nature Reviews Microbiology*, 15(10):591–605, October 2017.
- [17] Monica Rolando, Serena Sanulli, Christophe Rusniok, Laura Gomez-Valero, Clement Bertholet, Tobias Sahr, Raphael Margueron, and Carmen Buchrieser. *Legionella pneumophila* Effector RomA Uniquely Modifies Host Chromatin to Repress Gene Expression and Promote Intracellular Bacterial Replication. *Cell Host & Microbe*, 13(4):395–405, April 2013.
- [18] Pengfei Li, Dane Vassiliadis, Sze Ying Ong, Vicki Bennett-Wood, Chihiro Sugimoto, Junya Yamagishi, Elizabeth L. Hartland, and Shivani Pasricha. *Legionella pneumophila* Infection Rewires the *Acanthamoeba castellanii* Transcriptome, Highlighting a Class of Sirtuin Genes. *Frontiers in Cellular and Infection Microbiology*, 10:428, August 2020.
- [19] Sarah Rennie, Maria Dalby, Lucas van Duin, and Robin Andersson. Transcriptional decomposition reveals active chromatin architectures and cell specific regulatory interactions. *Nature Communications*, 9(1):487, December 2018.
- [20] Elphège P. Nora, Bryan R. Lajoie, Edda G. Schulz, Luca Giorgetti, Ikuhiro Okamoto, Nicolas Servant, Tristan Piolot, Nynke L. van Berkum, Johannes Meisig, John Sedat, Joost Gribnau, Emmanuel Barillot, Nils Blüthgen, Job Dekker, and Edith Heard. Spatial partitioning of the regulatory landscape of the X-inactivation centre. *Nature*, 485(7398):381–385, May 2012.
- [21] Robert J. Neff. Purification, Axenic Cultivation, and Description of a Soil Amoeba, *Acanthamoeba* sp. *The Journal of Protozoology*, 4(3):176–182, August 1957.
- [22] Rolf Michel and Bärbel Hauröder. Isolation of an *Acanthamoeba* Strain with Intracellular *Burkholderia pickettii* Infection. *Zentralblatt für Bakteriologie*, 285(4):541–557, April 1997.
- [23] David L. Rimm, Thomas D. Pollard, and Philip Hieter. Resolution of *Acanthamoeba castellanii* chromosomes by pulsed field gel electrophoresis and construction of the initial linkage map. *Chromosoma*, 97(3):219–223, November 1988.
- [24] Ritu Kundu, Joshua Casey, and Wing-Kin Sung. HyPo: Super Fast & Accurate Polisher for Long Read Genome Assemblies. preprint, Bioinformatics, December 2019.
- [25] Lyam Baudry, Nadège Guiglielmoni, Hervé Marie-Nelly, Alexandre Cormier, Martial Marbouth, Komlan Avia, Yann Loe Mie, Olivier Godfroy, Lieven Sterck, J. Mark Cock, Christophe Zimmer, Susana M. Coelho, and Romain Koszul. instaGRAAL: chromosome-level quality scaffolding of genomes using a proximity ligation-based scaffolder. *Genome Biology*, 21(1):148, June 2020.

- [26] Hervé Marie-Nelly, Martial Marbouty, Axel Cournac, Jean-François Flot, Gianni Liti, Dante Poggi Parodi, Sylvie Syan, Nancy Guillén, Antoine Margeot, Christophe Zimmer, and Romain Koszul. High-quality genome (re)assembly using chromosomal contact data. *Nature Communications*, 5(1):5695, December 2014.
- [27] Cyril Matthey-Doret, Lyam Baudry, Axel Breuer, Rémi Montagne, Nadège Guiglielmoni, Vittore Scolari, Etienne Jean, Arnaud Campreas, Philippe Henri Chanut, Edgar Oriol, Adrien Méot, Laurent Politis, Antoine Vigouroux, Pierrick Moreau, Romain Koszul, and Axel Cournac. Computer vision for pattern detection in chromosome contact maps. *Nature Communications*, 11(1):5795, November 2020. Number: 1 Publisher: Nature Publishing Group.
- [28] Tsung-Han S. Hsieh, Assaf Weiner, Bryan Lajoie, Job Dekker, Nir Friedman, and Oliver J. Rando. Mapping Nucleosome Resolution Chromosome Folding in Yeast by Micro-C. *Cell*, 162(1):108–119, July 2015.
- [29] Charlotte Cockram, Agnès Thierry, Aurore Gorlas, Roxane Lestini, and Romain Koszul. Euryarchaeal genomes are folded into SMC-dependent loops and domains, but lack transcription-mediated compartmentalization. *Molecular Cell*, 81(3):459–472.e10, February 2021.
- [30] Lyam Baudry, Gaël A Millot, Agnes Thierry, Romain Koszul, and Vittore F Scolari. Serpentine: a flexible 2D binning method for differential Hi-C analysis. *Bioinformatics*, 36(12):3645–3651, June 2020.
- [31] Leonid A. Mirny. The fractal globule as a model of chromatin architecture in the cell. *Chromosome Research*, 19(1):37–51, January 2011.
- [32] Tung B. K. Le, Maxim V. Imakaev, Leonid A. Mirny, and Michael T. Laub. High-Resolution Mapping of the Spatial Organization of a Bacterial Chromosome. *Science*, 342(6159):731–734, November 2013. Publisher: American Association for the Advancement of Science Section: Report.
- [33] Alaina Shumate and Steven L Salzberg. Liftoff: accurate mapping of gene annotations. *Bioinformatics*, 37(12):1639–1643, June 2021.
- [34] Thomas J. Byers. Molecular Biology of DNA in Acanthamoeba, Amoeba, Entamoeba, and Naegleria. In *International Review of Cytology*, volume 99, pages 311–341. Elsevier, 1986.
- [35] Claude Gicquaud and Nadine Tremblay. Observations with Hoechst Staining of Amitosis in *Acanthamoeba castellanii*. *The Journal of Protozoology*, 38(3):221–224, May 1991.
- [36] Qin Yang, Michael G. Zwick, and Marvin R. Paule. Sequence organization of the *Acanthamoeba* rRNA intergenic spacer: identification of transcriptional enhancers. *Nucleic Acids Research*, 22(22):4798–4805, 1994.
- [37] C Rabl. Über zelltheilung. *Morphol. Jahrbuch*, (10):214–330, 1885.
- [38] Luce Mengue, Matthieu Régnaq, Willy Aucher, Emilie Portier, Yann Héchard, and Ascel Samba-Louaka. Legionella pneumophila prevents proliferation of its natural host *Acanthamoeba castellanii*. *Scientific Reports*, 6(1):36448, December 2016.
- [39] Jonay Garcia-Luis, Luciana Lazar-Stefanita, Pilar Gutierrez-Escribano, Agnes Thierry, Axel Cournac, Alicia García, Sara González, Mar Sánchez, Adam Jarmuz, Alex Montoya, Marian Dore, Holger Kramer, Mohammad Mehdi Karimi, Francisco Antequera, Romain Koszul, and Luis Aragon. FACT mediates cohesin function on chromatin. *Nature structural & molecular biology*, 26(10):970–979, August 2019.
- [40] Martial Marbouty, Antoine Le Gall, Diego I. Cattoni, Axel Cournac, Alan Koh, Jean-Bernard Fiche, Julien Mozziconacci, Heath Murray, Romain Koszul, and Marcelo Nollmann. Condensin- and Replication-Mediated Bacterial Chromosome Folding and Origin Condensation Revealed by Hi-C and Super-resolution Imaging. *Molecular Cell*, 59(4):588–602, August 2015.
- [41] Anna Anchimiuk, Virginia S Lioy, Florian Patrick Bock, Anita Minnen, Frederic Boccard, and Stephan Gruber. A low Smc flux avoids collisions and facilitates chromosome organization in *Bacillus subtilis*. *eLife*, 10:e65467, August 2021. Publisher: eLife Sciences Publications, Ltd.
- [42] Fidel Ramírez, Vivek Bhardwaj, Laura Arrigoni, Kin Chung Lam, Björn A. Grüning, José Villaveces, Bianca Habermann, Asifa Akhtar, and Thomas Manke. High-resolution TADs reveal DNA sequences underlying genome organization in flies. *Nature Communications*, 9(1):189, January 2018. Number: 1 Publisher: Nature Publishing Group.
- [43] Li Song and Liliana Florea. Rcorrector: efficient and accurate error correction for Illumina RNA-seq reads. *GigaScience*, 4(1):1–8, December 2015. Number: 1 Publisher: BioMed Central.
- [44] Maxim Imakaev, Geoffrey Fudenberg, Rachel Patton McCord, Natalia Naumova, Anton Goloborodko, Bryan R. Lajoie, Job Dekker, and Leonid A. Mirny. Iterative correction of Hi-C data reveals hallmarks of chromosome organization. *Nature Methods*, 9(10):999–1003, October 2012. Number: 10 Publisher: Nature Publishing Group.

- [45] Tao Yang, Feipeng Zhang, Galip Gürkan Yardımcı, Fan Song, Ross C Hardison, William Stafford, Feng Yue, and Qunhua Li. HiCRep: assessing the reproducibility of Hi-C data using a stratum-adjusted correlation coefficient. page 37.

6 Supplementary figures

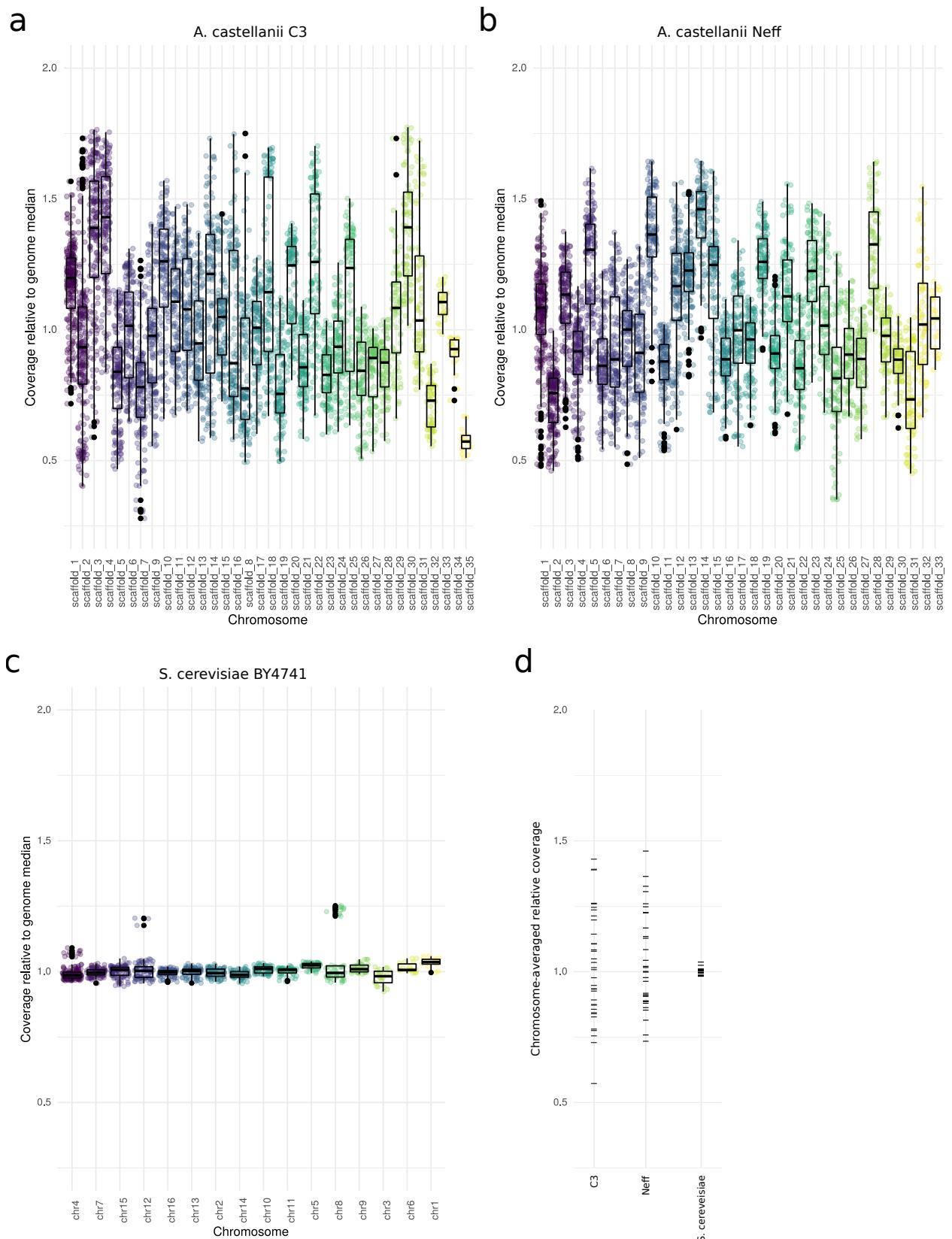


Figure S1: Coverage across scaffolds of *A. castellanii* compared to a known haploid. Distribution of mean Illumina shotgun coverage in 100kb sliding windows, normalized by the genome median, across *A. castellanii* Neff (a) and C3 (b) scaffolds compared to *Saccharomyces cerevisiae* strain BY4741, a known haploid (c). Variability of median coverage per chromosome for all three organisms. For *S. cerevisiae*, library SRR1569870 was used.

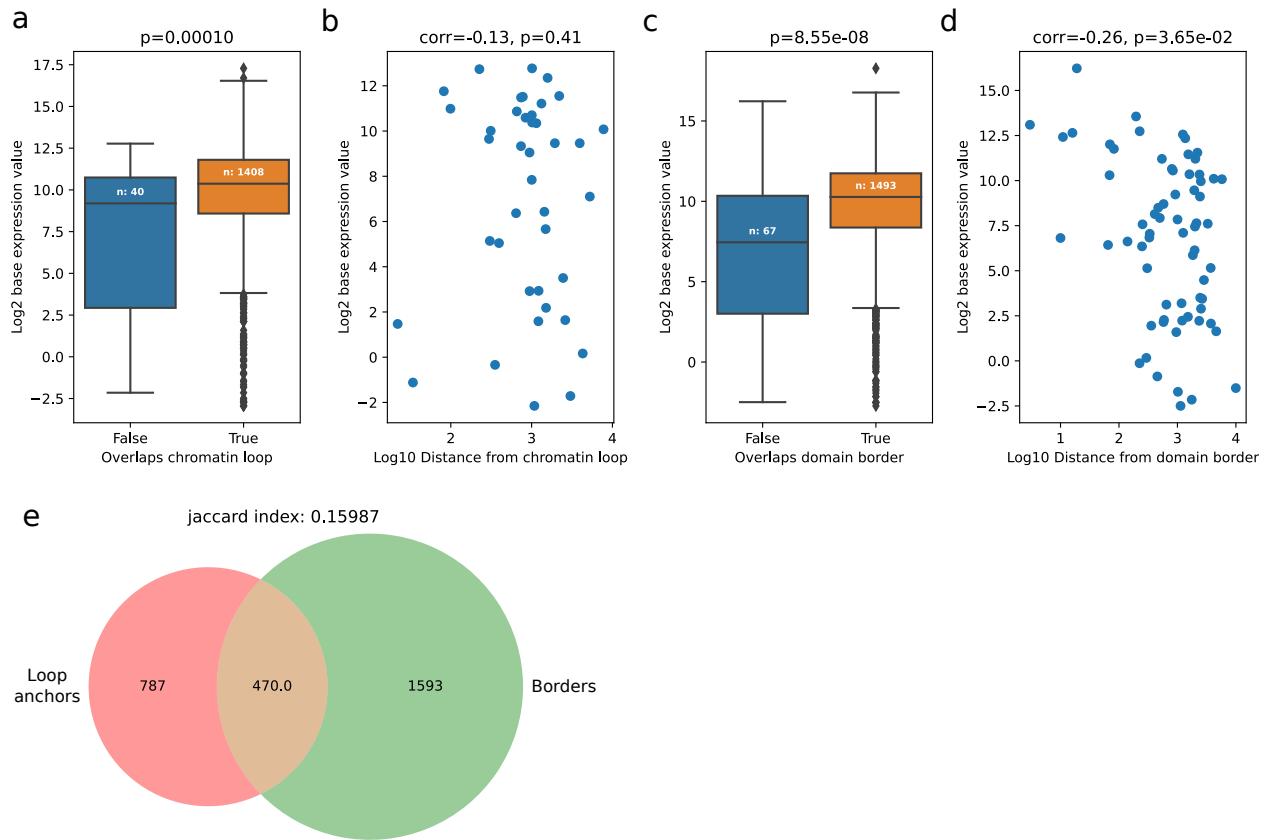


Figure S2: Gene expression according to position relative to chromatin structures. Gene expression versus (a) overlap status with chromatin loops and (b) distance to closest loop. Gene expression versus (c) overlap status with domain borders and (d) distance to closest border. P-values reported for overlap comparisons are obtained using Mann-Whitney U test, correlation coefficients and associated p-values are computed using Spearman's correlation test. (e) Overlap between chromatin loop anchors and domain borders represented as a Venn diagram.

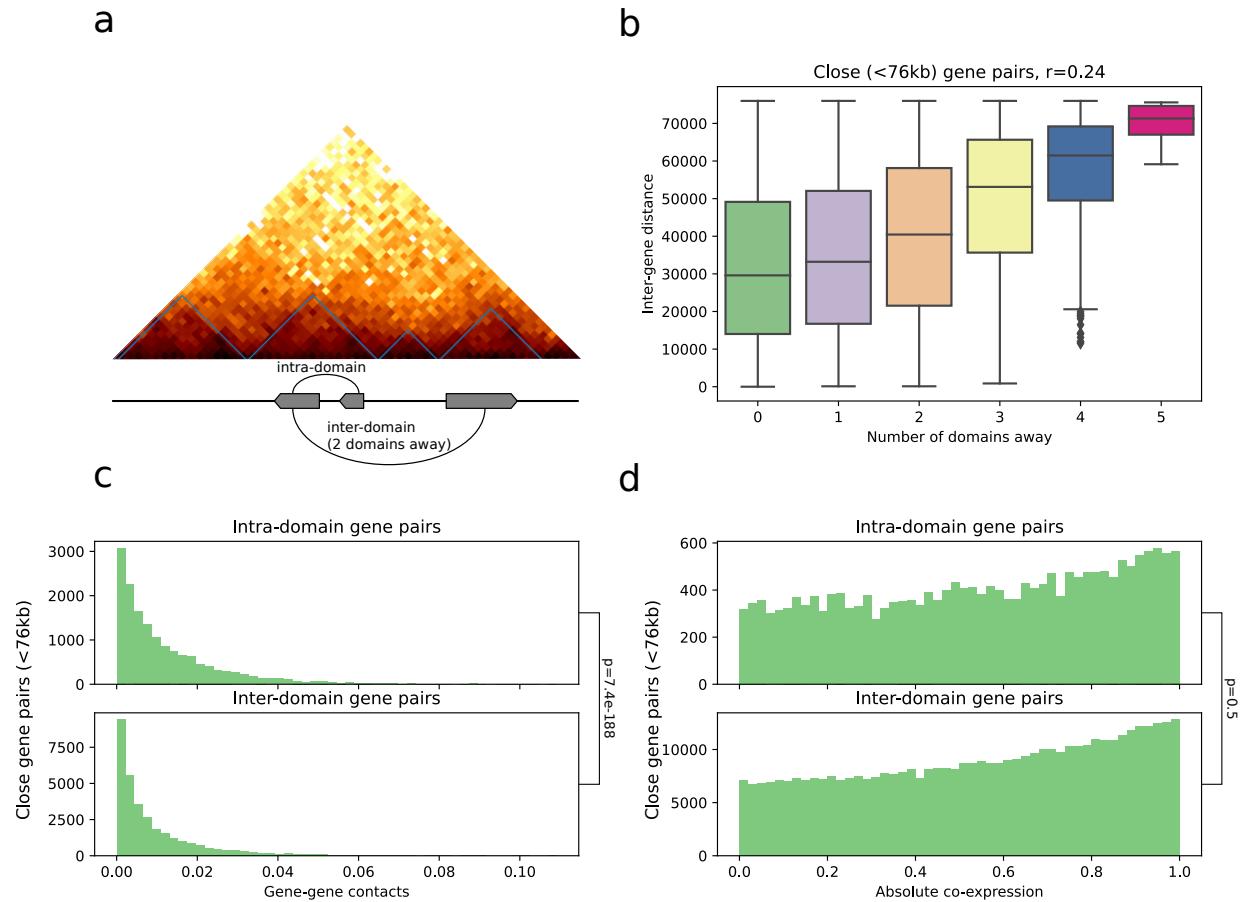


Figure S3: Relationship between genes and insulation domains. (a) Example domains detected by Chromosight in the C3 strain, with theoretical genes for demonstration. (b) Relationship between inter-gene distance and number of domains separating them. (c) Distribution of mean inter-gene contacts according to domain separation status. (d) Distribution of gene-pairs co-expression according to domain separation status. For all panels, only gene pairs separated by less than the median domain size were selected.

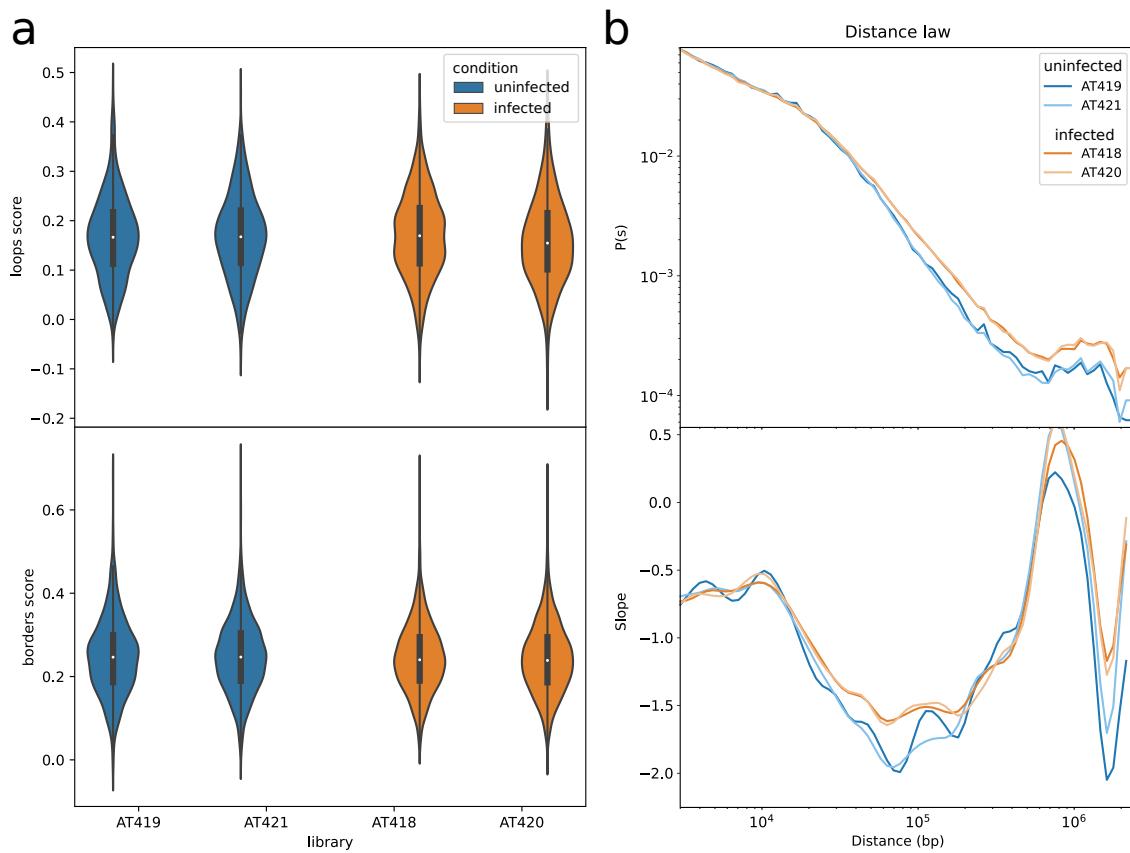


Figure S4: Global comparisons of infection Hi-C results between replicates. (a) Distribution of Chromosight loops and borders scores for all 4 samples. (b) Distance-contact decay function (denoted $P(s)$) and its slope.

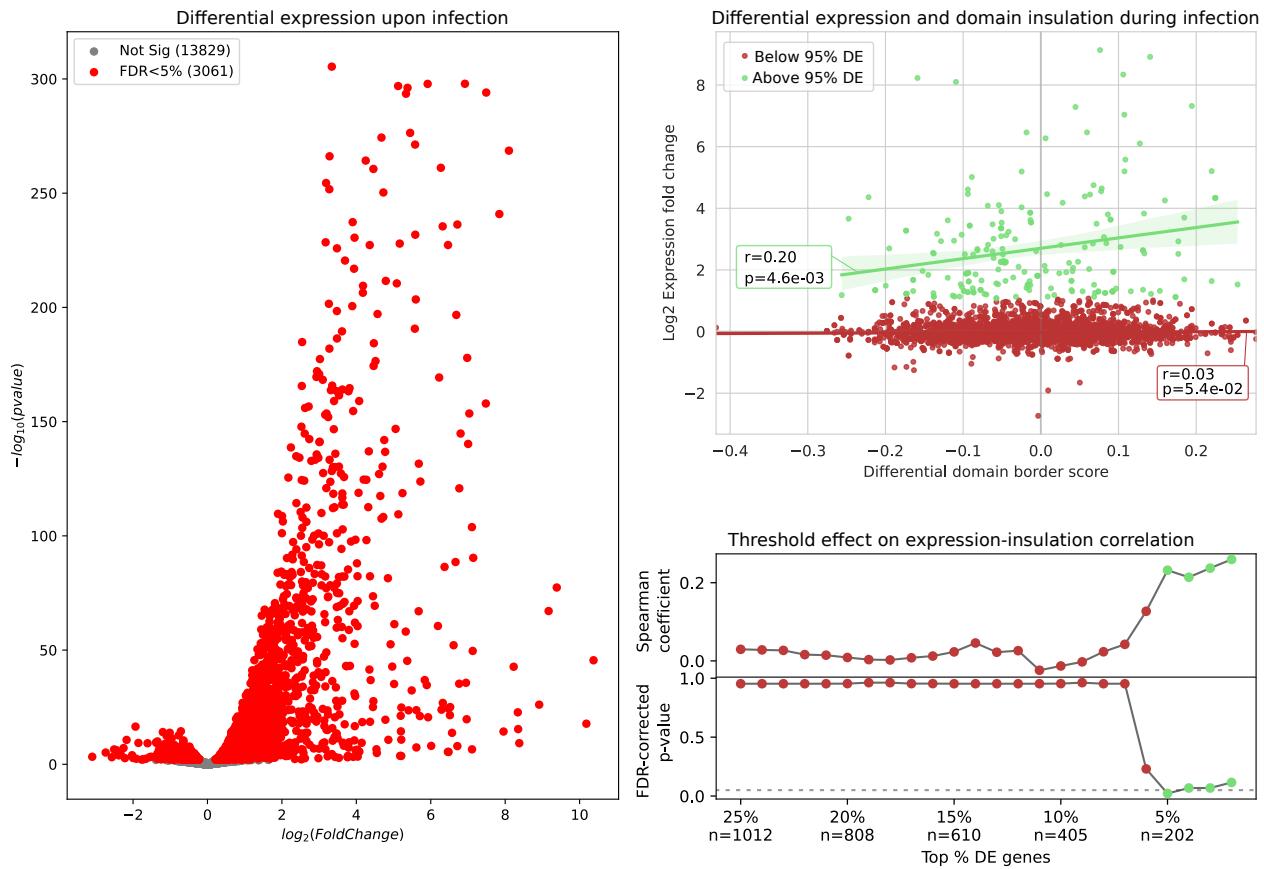


Figure S5: Relationship between differential expression and domain insulation during infection. (a) Volcano plot showing differential gene expression (DE) of infected (5h p.i.) versus uninfected amoeba. Genes with significant corrected p-values (FDR<5%) are shown in red. (b) Changes in gene expression and insulation strength of closest domain border during infection. Linear regression lines, Spearman correlation coefficients and associated p-values are shown separately for genes with extreme fold change values (95% quantile) and the rest. (c) Spearman correlation coefficient between expression fold change and domain insulation change, and associated FDR-corrected p-values (FDR<5%) for different subsets of genes according to the threshold of extreme fold change. Values are colored according to the 95% threshold selected in b.

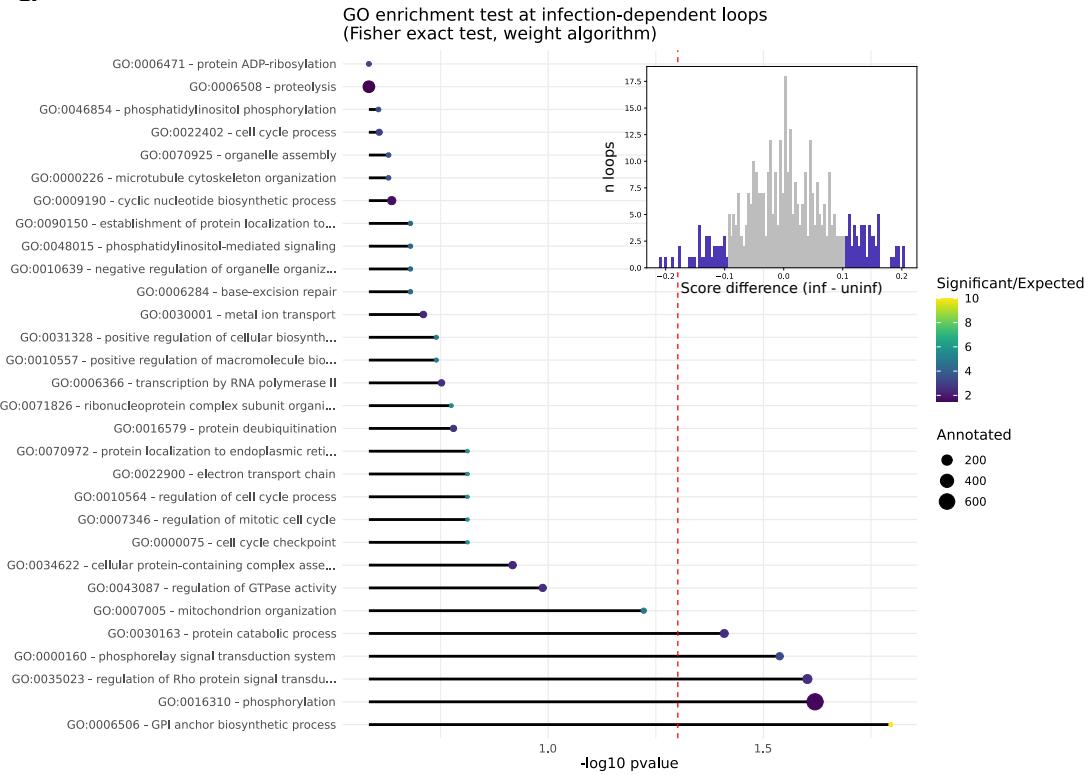
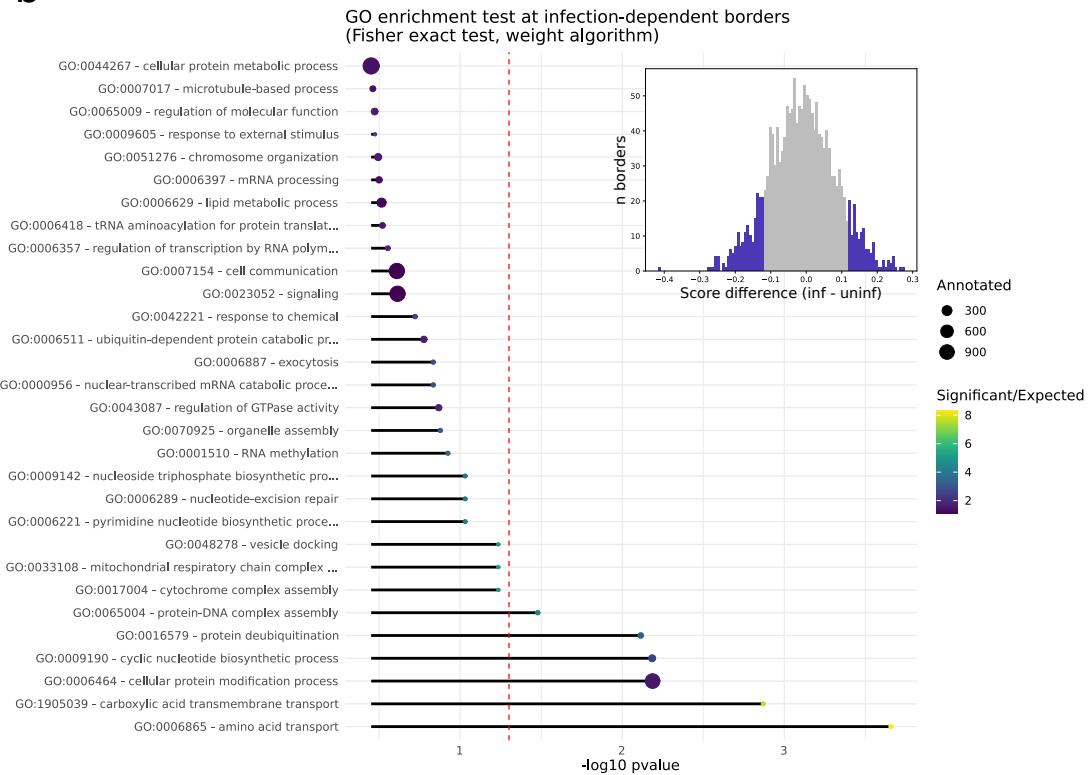
a**b**

Figure S6: GO term enrichment test results for genes overlapping infection-dependent chromatin loops (a) and domain borders (b). Histograms show the distribution of loop and border score changes during infection, with highlighted portions showing the 80% percentile threshold used to include genes in the GO enrichment test.

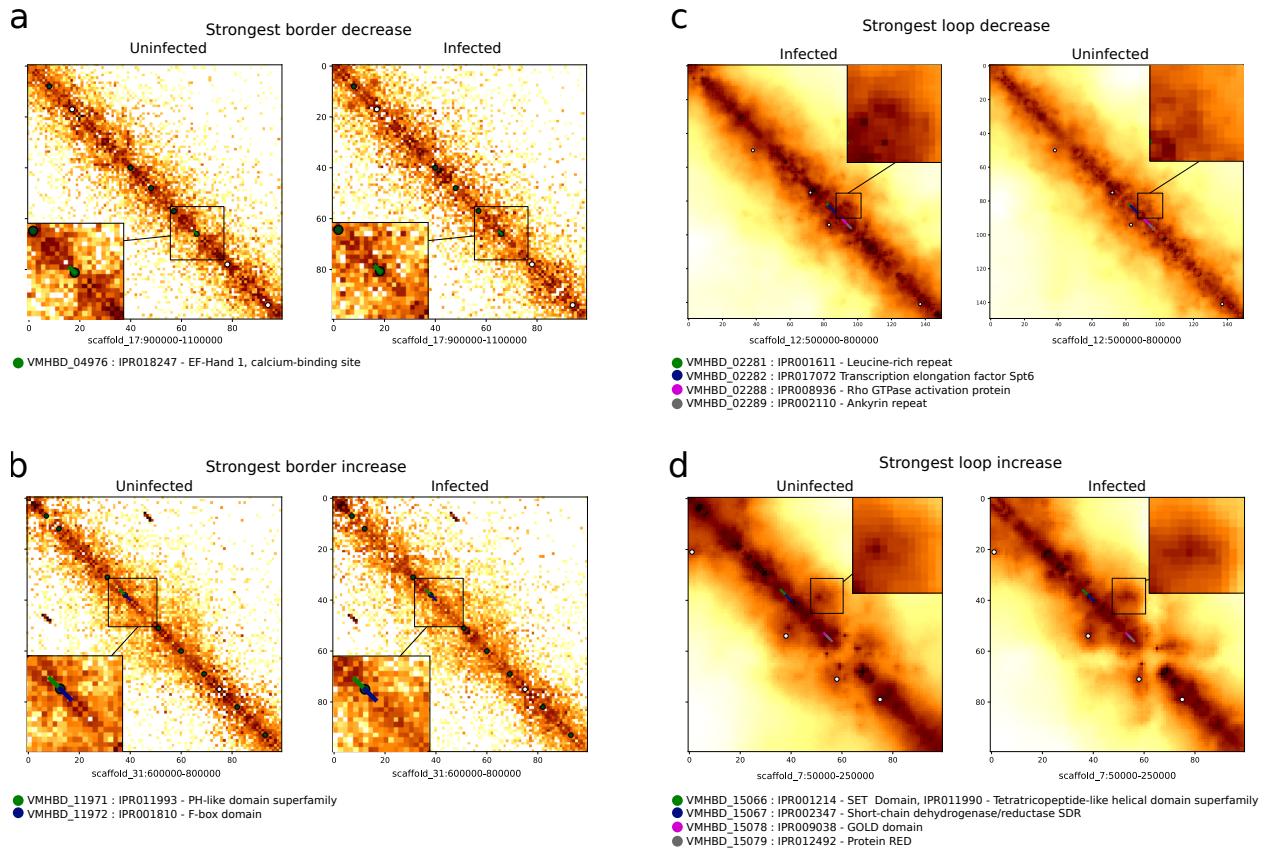


Figure S7: Hi-C zooms on strongest pattern changes during infection. Description of the closest genes are shown below each zoom. Balanced contact map zooms showing strongest border decrease (a) and increase (b). Serpentine-binned contact maps showing strongest loop decrease (c) and increase (d).