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

A PREPRINT

August 4, 2021

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. Infection of *A. castellanii* strain C3 induced heightened chromatin loop formation and increased interchromosomal contacts and global telomere clustering.

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, where the bacterium replicates 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. acanthamoeba* C3 strain by *L. pneumophila* through the prism of the three-dimensional organisation of the amoeba genome. This analysis required the generation of a good 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. xx 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*. xx

2 Results

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].

2.1 Genome assembly

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, 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

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.

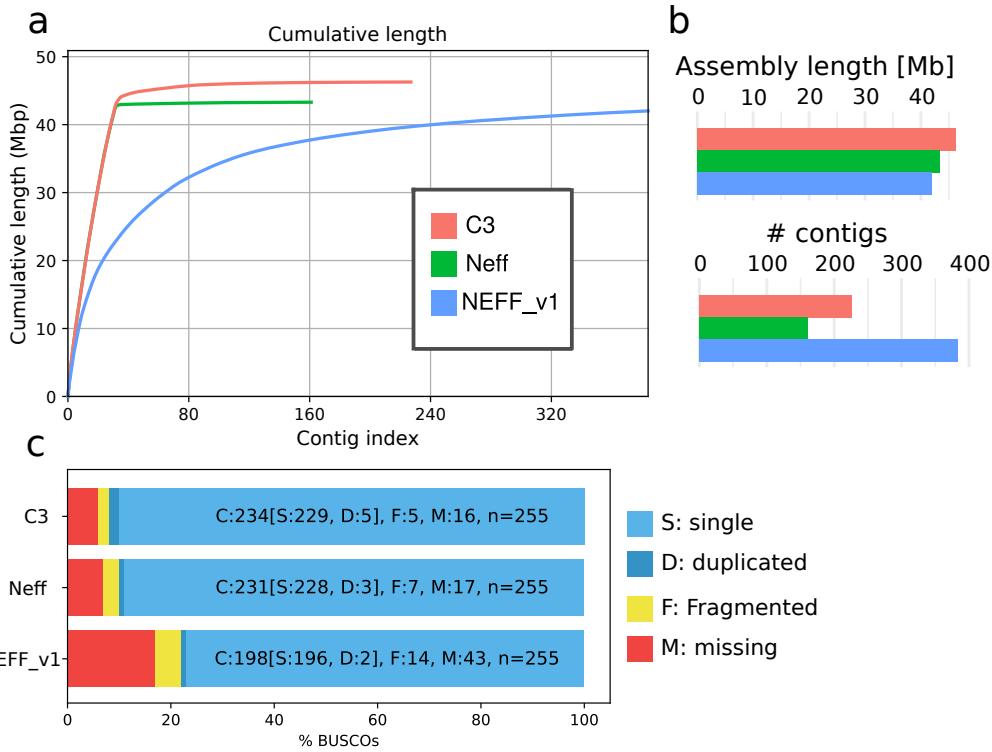


Figure 1: **Assembly statistics for *A. castellanii*.** Comparison genome assemblies for strains C3 and Neff, versus the previous NEFF-v1 genome assembly [3]. **a**, Cumulative length plot showing the relationship between number of contigs (largest to smallest) and lenght of the assembly. **b**, General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

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 misassemblies in genome sequences ([26].

2.2 Comparison of Neff and C3 strains

General features of the Neff and C3 genomes are largely conserved. The genome size is approximately the same, karyotypes are very similar and aligned segments have a substitution probability of 6.9%.

Clarke et al. reported over a hundred putative HGT events in the sequence of *A. castellanii* [3]. We aligned the sequences of those genes to both strains' assemblies and retrieved 172 and 117 matching genes in Neff and C3, respectively. We found no difference in sequence composition (GC content, GC skew, 2-mer frequency profile) between those HGT and the rest of genes in either strain. We found that HGT also had similar length and number of exons from other genes in Neff, however the matched HGT in C3 were shorter and had fewer exons than the rest of the genome.

2.3 Viral sequences

Several species of giant viruses are associated to *A. castellanii* and have integrated into its genomic sequence [3]. As viruses are often considered horizontal gene transfer agents, we investigated whether the integrated viral sequences are neighbouring horizontally transferred genes reported by Clarke et al. We identified multiple viral insertions of up to 2.5kb from Pandoravirus and Aureococcus, but we found no particular association between the chromosomal positions of bacterial HGT and viral sequences.

2.4 Spatial organisation of the *A. castellanii* genome

Visualising the Hi-C contact maps of both strains shows the strong concordance between *A. castellanii* chromosomes and our assembly (**Fig. 2**). In Neff, the highest intensity contacts are all on the diagonal, suggesting no large-scale misassembly. The C3 assembly retains a few misassembled blocks, especially in the rDNA region where tandem repeats could not be resolved correctly. The contact maps of these chromosome-resolved assemblies reveal a grid-like pattern with higher intensity dots towards the telomeres. As illustrated on the right of (**Fig. 2**), these contacts can be interpreted as a clustering of the telomeres from different chromosomes. This is reminiscent - but different - from the Rabl conformation in yeast, where all centromeres cluster to the spindle-pole body. Additionally, 5S ribosomal DNA genes, although dispersed throughout the genome, are enriched towards telomeres (Fig. 2). All 3 rDNA genes (18S, 28S and 5S) are in close spatial proximity to telomeres.

Although 5S ribosomal RNA genes are dispersed throughout the *A. castellanii* genome, Hi-C data suggests they cluster in space. It is estimated that *A. castellanii* has 24 copies of the 5S rDNA gene, but each cell contains around 600 copies because of polyploidy [27].

Chromosomal contact maps of *A. castellanii* show typical features of eukaryotic genomes, such as chromatin loops and insulation domains. Most chromatin loops are regularly spaced and have a typical size of 20kb, which is similar to mitotic loops observed in *S. cerevisiae*.

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

We investigated the effect of *L. pneumophila* infection on genomic contacts and transcription of *A. castellanii* str. C3. As the bacterium reorganizes the host transcriptional program, we started by looking at large scale changes. We found that interchromosomal contacts are globally altered during infection. Chromosomes bearing 18S and 28S rDNA genes show stronger interactions with other chromosomes during infection. Although inter-chromosomal contact changes are highly heterogenous across chromosomes, there is a global increase in the strength of inter-telomeric clustering. Chromatin loops are globally heightened, and regularly spaced.

It was previously reported that *Legionella* infection reduces proliferation in *A. castellanii* [?]. The large scale 3D changes we observe are reminiscent of cell cycle block in yeast and would suggest that the bacterium stops its host's cell cycle at a specific checkpoint. It was previously shown that *L. pneumophila* infection interferes with the host cell cycle and is associated with a decrease of CDC2b mRNA [28].

More recently, Li et al. [18] showed that Legionella infection of *A. castellanii* causes major changes in the transcriptome and disrupts the expression of many cell-cycle-related genes. We reused their differential expression results to compute the correlation between gene co-expression and inter-gene contacts. Unlike previous results in mammals, we found no increase in co-expression for genes sharing the same contact domain. This suggests that chromatin structure in *A. castellanii* may have a lesser importance in dictating gene expression programs than in mammals.

We used CHESS to detect structural changes happening during infection and found one region which was consistently different between both replicates. This region contained several genes including ankyrin repeat and leucin-rich repeat containing proteins, and small GTP-ases.

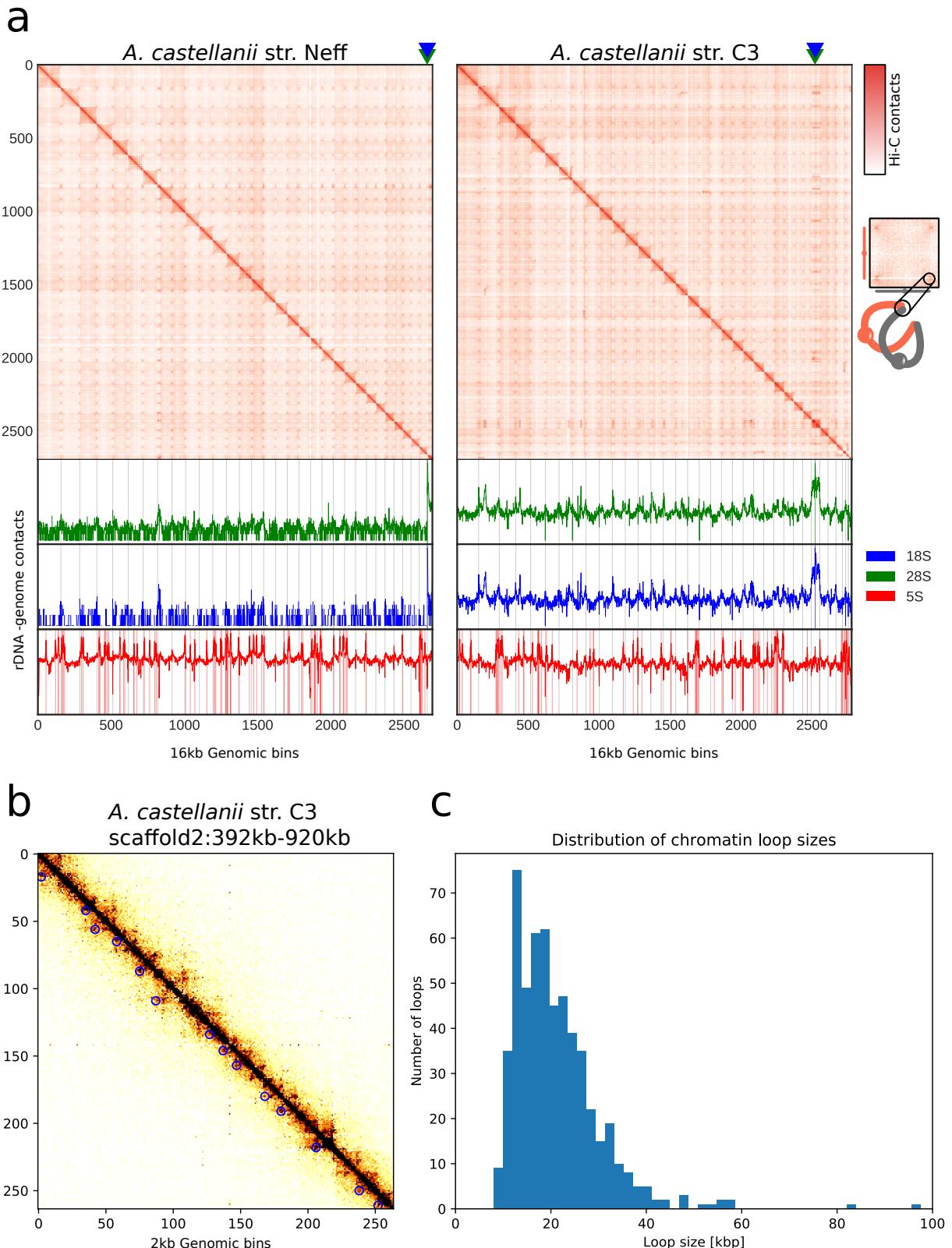


Figure 2: **Spatial organisation of the *A. castellanii* genome.** Top: Hi-C contact maps of the Neff (left) and C3 (right) strains. The genome is divided into 16kb bins and each pixel represents the contact intensity between a pair of bins. Each scaffold is visible as a red square on the diagonal. In 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 separated 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 5S rDNA sequences are shown with vertical red lines on the bottom panel.

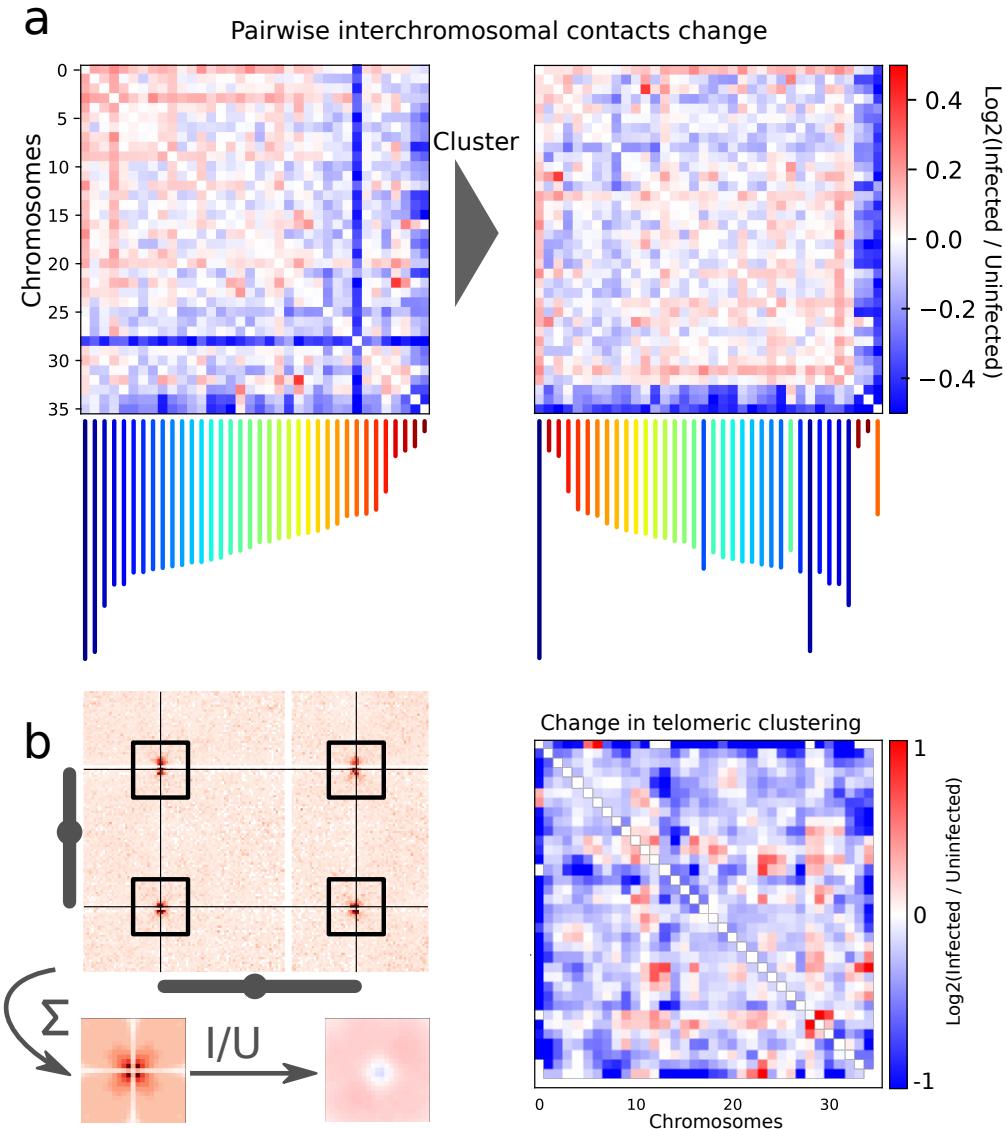


Figure 3: **Interchromosomal contacts change during *L. pneumophila* infection.** **a**, Average contact change during infection between each pair of chromosomes before (left) and after (right) clustering chromosomes by interchromosomal contact profile. **b**, Inter-telomeric contacts between pairs of 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 chromosome, consisting of the average of 4 telomeric windows. Intrachromosomal windows are excluded.

3 Discussion

Considering the biology of *A. castellanii*, with suspected amitosis and probable aneuploidies, a dispersion of 5S ribosomal DNA across all chromosomes would make sense, to ensure a consistant copy number of 5S rDNA in daughter cells. The increase in interchromosomal contacts with rDNA-containing chromosomes could reflect a global increase in translational activity during infection. The increased production of ribosomes could be linked to alterations in the nucleolus structure.

4 Methods

4.1 Bacterial infections

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.

4.2 Library preparations

4.2.1 Hi-C

Hi-C libraries were prep arima, read length, Illumina machine 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.

4.2.2 Shotgun

Kit / protocol, read length, Illumina machine Shotgun libraries were sequenced by Novogene at 2x150bp on a Novaseq machine.

4.2.3 RNA-seq

Company, protocol, machine

4.2.4 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.

4.3 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.

Minimap2 v2.17 was used for all long reads alignments, and bowtie2 v2.3.4.1 for short reads alignments. assembly snakemake pipeline, hypo, flye, instagraal, manual processing, funannotate et cie.

The pipeline for genome assembly is available on Github at: ...

4.4 Genome annotation

The Neff and C3 assemblies were annotated via funannotate v1.5.3. Repeated sequences were masked using repeatmasker. For gene prediction, RNA-seq reads from Neff were aligned to both genomes using STAR v2.7.0. The resulting

alignments were used to train AUGUSTUS v3.3.1 and BRAKER via funannotate predict. Predicted proteins were fed to Interproscan v5.22, Phobius v1.7.1 and EggNOG-mapper v2.0.0 were used to generate functional annotations.

The genome annotation pipeline is available on Github at: ...

Ribosomal DNA genes were annotated separately using RNAmmer v1.2 with HMMER 2.3.2.

4.5 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.

4.6 Hi-C analyses

Infection snakemake pipeline, hicstuff, cooler, chromosight, pareidolia Reads were aligned with bowtie2 v2.4.1 and Hi-C matrices were generated using hicstuff v2.3.0. 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 [29]. Loops and domain borders were detected and quantified using Chromosight.

Detection of changes in pattern intensities were performed using pareidolia (<https://www.github.com/koszullab/pareidolia>). Pareidolia uses a median-background change detection algorithm on Chromosight's pattern correlation maps.

Pairwise sample similarities from Hi-C matrices were computed using hicreppy v0.0.6 <https://www.github.com/cmdoret/hicreppy>, which uses the HiCRep algorithm [30].

The pipeline for Hi-C analysis of *A. castellanii* infections is available on github at: ...

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] Heng Li, Xiaowen Feng, and Chong Chu. The design and construction of reference pangenome graphs with minigraph. *Genome Biology*, 21(1):265, December 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 Guigielmoni, Hervé Marie-Nelly, Alexandre Cormier, Martial Marbouty, 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] 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.
- [28] Luce Mengue, Matthieu Régnacq, 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.
- [29] 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.
- [30] 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.