Thesis for the degree

Master of Science

Submitted to the Scientific Council of the

Weizmann Institute of Science

Rehovot, Israel

By

Meir Yehihel Sylman

עבודת גמר (תזה) לתואר

מוסמך למדעים

מוגשת למועצה המדעית של

מכון ויצמן למדע

רחובות, ישראל

מאת

מאיר יחיאל סילמן

חקר ויסות ביטוי גנים בלישמניה

ברמת יחידות תעתוק פוליציסטרוניות

Investigating Gene Expression Regulation in *Leishmania* at the Level of Polycistronic Transcription Units

Advisor:

Prof. Yitzhak Pilpel

מנחה:

פרופסור יצחק פלפל

January 2026

שבט תשפ"ו

Table of Contents

Abstract

Leishmania are protozoan parasites that alternate between two hosts: a sandfly vector, where they proliferate as motile promastigotes, and a mammalian host, where they persist inside phagocytic cells as amastigotes in the phagolysosome. These transitions require efficient remodeling of gene expression. Unlike most eukaryotes, Leishmania genes are organized into large polycistronic transcription units (PTUs) that appear to be transcribed without canonical promoters, so regulation is commonly attributed to chromosome copy-number alterations and post-transcriptional control. This thesis tests whether PTUs provide an intermediate regulatory layer. I found that despite common notion in the field, genes that belong to the same PTU may change expression levels coherently when environmental conditions change, while sometimes being distinct from expression changes of neighboring PTU on shared chromosome. A search for promoter-like motifs offered limited support for transcription factor based control; instead, DNA physical properties deduced from sequence composition emerged as a plausible contributor. I found that GC-rich genes at PTU beginnings show lower basal expression but stronger upregulation under heat stress, consistent with temperature relieving thermodynamic constraints on transcription. In parallel, I examined translational tuning through adaptation between codon usage and the cellular tRNA pool. Across trypanosomatids, integrating codon usage, inferred tRNA pools, and the tRNA adaptation index (tAI) shows that translation-efficiency profiles track evolutionary relationships and reveal lineage- and function-specific optimization. I found that a specific pattern of chromosome aneuploidy, that occurs in 37°C, increased DNA copy number of tRNA genes whose corresponding codons are enriched in an amastigote-related gene. Together, the results support a model in which PTU architecture and translation efficiency jointly shape adaptive gene expression.

List of abbreviations

BLAST- Basic Local Alignment Search Tool

CDS – Coding Sequence

ChIP-seq - Chromatin Immunoprecipitation followed by Sequencing

CSS - Convergent Strand switch

DSS - Divergent Strand switch

GPI - Glycosylphosphatidylinositol

L. – Leishmania

MEME - Multiple Expectation maximizations for Motif Elicitation

PCA- Principle component analysis

PRO-seq -

PTU- Polycistronic Transcription Unit

RPKM - Reads Per Kilobase Million

T. – Trypanosoma

tAI – tRNA Adaptation Index

TPM- Transcripts Per Million

Introduction

Leishmania's Life Cycle and Environmental Challenges

Leishmania is a protozoan parasite belonging to the class Kinetoplastea within the family Trypanosomatidae[1]. More than 20 *Leishmania* species are known to cause disease in humans. Among them, *Leishmania donovani* and *Leishmania major* are two of the most clinically significant species, responsible for visceral and cutaneous leishmaniasis, respectively[2–4]. Visceral leishmaniasis is characterized by infection of internal organs such as the liver, spleen, and bone marrow and is fatal if left untreated, with an estimated annual incidence of 50,000–90,000 cases. Cutaneous leishmaniasis, caused primarily by *L. major*, accounts for approximately 230,000–430,000 cases per year worldwide[1–3]. Leishmaniasis is endemic across large regions of Asia, Africa, the Americas, and southern Europe[1,4,5]. Transmission can be anthroponotic in some regions[6,7], occurring directly between humans, while in others it is zoonotic, involving animal reservoirs such as domestic dogs[8–10].

The *Leishmania* life cycle alternates between two distinct hosts and cellular forms (**Figure 1A**). In the insect vector, typically sandflies of the genus *Phlebotomus*, the parasite exists as extracellular promastigotes, which are elongated and flagellated. Yet in the insect vector, the parasites differentiate into the metacyclic promastigote form, which is infective to mammals. Upon transmission to the mammalian host, parasites are phagocytosed by macrophages and differentiate into intracellular amastigotes, which are round and non-flagellated.

This transition exposes the parasite to profound environmental changes. In the sandfly midgut, nutrient availability progressively declines following the blood meal. In contrast, within mammalian host macrophages, parasites encounter elevated temperature, acidic pH, and immune pressure. Successful infection therefore requires extensive and rapid adaptation of parasite physiology and gene expression.

The Unique Genome Architecture of *Leishmania*

The *Leishmania* genome exhibits several unusual structural features that distinguish it from most eukaryotic model organisms. Chromosome copy number is not strictly diploid; instead, *Leishmania* displays constitutive and condition-dependent aneuploidy[11,12]. While the median chromosome copy number is typically close to two, several chromosomes frequently appear in three copies, and

chromosome 31 is consistently found as tetraploid[13]. Importantly, patterns of aneuploidy can vary between species and in response to environmental conditions.

Approximately 8,000 protein-coding genes are encoded in the *Leishmania* genome. Unlike most eukaryotes, these genes are not individually regulated by gene-specific promoters. Instead, they are organized into large Polycistronic Transcription Units (PTUs), in which typically tens of genes, mostly with unrelated function, share the same orientation are co-transcribed by RNA polymerase II[14]. Transcription initiates at the beginning of a PTU and proceeds continuously until it terminates at its end, or prematurely. The orientation of adjacent PTUs can be divergent strand-switch (dSS) when the two clusters are orientated head-to-head in 5' to 3' manner, convergent strand-switch (cSS) when they oriented tail to tail in 5' to 3' manner, or tandem. Another key feature of transcriptional organization in kinetoplastids is the presence of the modified DNA base, β-D-glucosyl-hydroxymethyluracil, termed base J. Base J is found at transcription termination sites[15] and serves as a conserved signal marking the ends of PTUs. An exception to the clustered organization of the genes is the spliced leader RNA genes, which possess individual promoters[16].

Following polycistronic transcription, precursor RNAs undergo trans-splicing, during which a capped spliced leader sequence is added to the 5′ end of each mRNA, while polyadenylation occurs at the 3′ end. As a result, mature mRNAs are generated post-transcriptionally from long polycistronic precursors[17].

In addition, *Leishmania* genomes contain a relatively low number of tRNA genes compared to other eukaryotes. These tRNA genes are often arranged in clusters and are unevenly distributed across chromosomes. Notably, several tRNA gene clusters are located at divergent strand-switch regions and have been suggested to contribute to chromosome organization and possibly centromere-associated functions[18].

Trypanosomatids: Evolutionary Context

While this study focuses primarily on *Leishmania*, several analyses also include related trypanosomatid species to enable comparative and evolutionary interpretations. The Trypanosomatidae family comprises parasitic and free-living protists with diverse life cycles and host associations. We made phylogenetic analyses[19] (**Figure 1C**) of this group, inferred using maximum-

likelihood approaches, and revealed the evolutionary relationships among *Leishmania*, *Trypanosoma*, and related taxa.

Including multiple species allows the identification of conserved and lineage-specific features of genome organization and gene regulation. Such comparisons provide a broader framework for understanding whether observed regulatory patterns are unique to *Leishmania* or represent more general properties of trypanosomatid biology.

The Known Gene Expression Regulation Mechanisms

The polycistronic organization of protein-coding genes implies an apparent absence of transcription initiation regulation at the level of individual genes. Indeed, no canonical promoter motifs or gene-specific transcription factors have been identified for most *Leishmania* genes. At the PTU level, transcription by RNA polymerase II is also unusual, as initiation sites lack well-defined sequence features typically associated with eukaryotic promoters. Regulation of gene expression in *Leishmania* is generally thought to occur predominantly through aneuploidy and post-transcriptional mechanisms[20]. At the DNA level, gene expression can be modulated by changes in chromosome copy number and by individual or clustered gene copy number variations[21,22]. At the RNA level, differences in mRNA abundance arise through regulated mRNA stabilization and degradation[23,24].

PTU as an expression regulatory element

Current models of gene regulation in *Leishmania* emphasize two primary hierarchical levels: chromosome-wide regulation through aneuploidy and gene-level regulation via post-transcriptional mechanisms. However, this framework leaves open the question of whether intermediate regulatory units of gene expression regulation exist.

Given that PTUs represent the fundamental transcriptional units of the genome, it is plausible that regulation may occur at the level of individual PTUs, independently of both whole-chromosome effects and the gene-specific processes. Our hypothesis suggests that if PTUs are functioning as independent regulatory units, the pattern of their transcriptional responses to environmental stress and adaptations should differ from those of neighboring PTUs and from the general trend of the response of the chromosome on which they reside. Conversely, if observed expression changes are driven solely by aneuploidy, PTUs would be expected to follow chromosome-wide trends. Alternatively, if regulation operates exclusively at the individual gene level, we expect not to detect

independence at the level of the PTU, comparing it to its genomic neighborhood. And if observed expression changes are driven by both chromosome and individual gene levels, we aren't expecting to see a uniqueness in the genes located in shared PTU.

This study investigates whether PTUs in *Leishmania* exhibit independent regulatory behavior, constituting an additional layer of gene expression regulation. To address this question, transcriptomic data were analyzed from a laboratory evolution experiment in which *Leishmania donovani* parasites were exposed to prolonged environmental and chemical stresses, including elevated temperature, low pH, nutrient limitation, and drug treatment (**Figure 1B**).

In this experimental system, parasites were cultured over multiple passages under defined stress conditions, with untreated controls maintained in parallel. Samples were collected at selected time points for RNA sequencing and DNA sequencing, enabling the joint analysis of transcriptional responses and chromosome copy number changes.

By comparing PTU-specific expression responses to those of adjacent genomic regions and entire chromosomes, this study aims to determine whether PTUs behave as autonomous regulatory units. In addition, the work explores the contribution of translation efficiency- through codon usage, tRNA gene content, and tRNA adaptation index- to condition specific and species specific adaptive gene expression regulation. Together, these analyses seek to refine the current understanding of gene regulation in *Leishmania*.

Figure 1

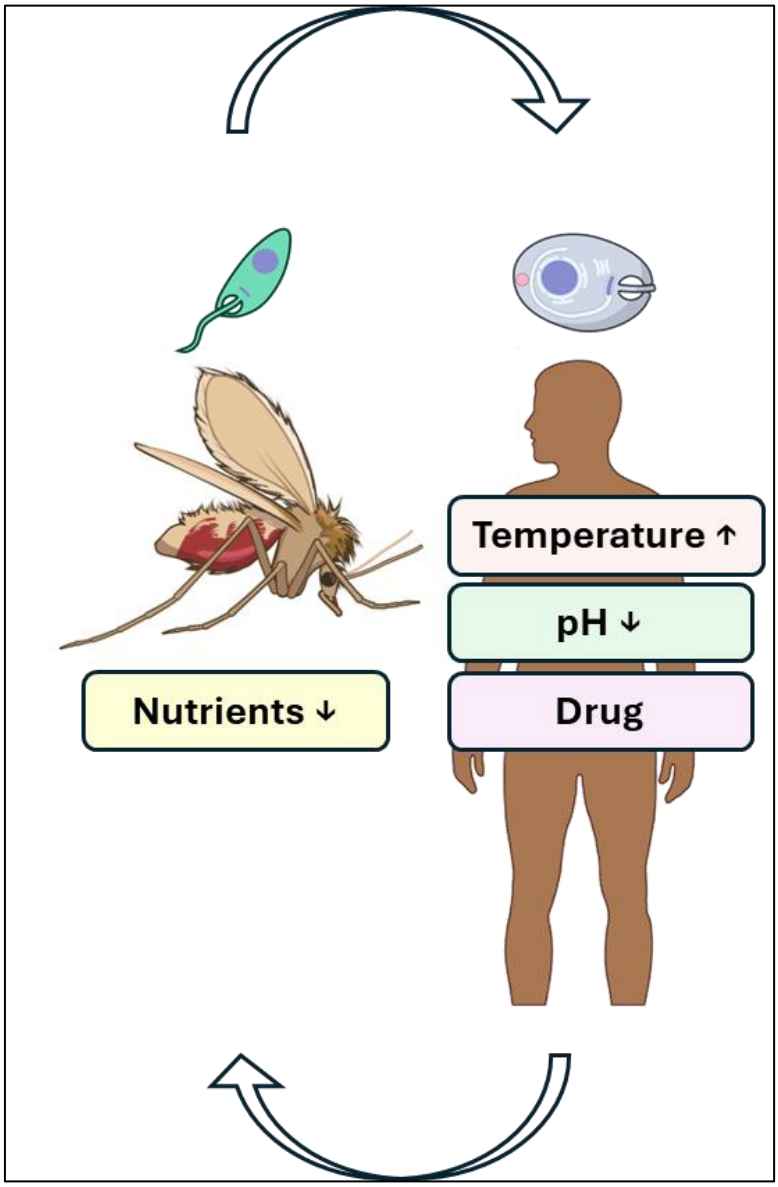


Figure 1A: *Leishmania* life cycle. *Leishmania* alternates between a mammalian host and a phlebotomine sand fly vector. In the mammalian host, parasites are taken up by macrophages and differentiate into amastigotes, a rounded, non-flagellated intracellular form also within the acidic phagolysosomal compartment. This stage experiences host body temperature and, during treatment, may be exposed to anti-leishmanial drugs. During a blood meal, infected sand flies ingest amastigotes, which transform in the fly midgut into flagellated promastigotes (elongated, extracellular). In the sand fly, the parasite develops in a changing midgut environment, including progressive depletion of nutrients as the blood meal is digested.

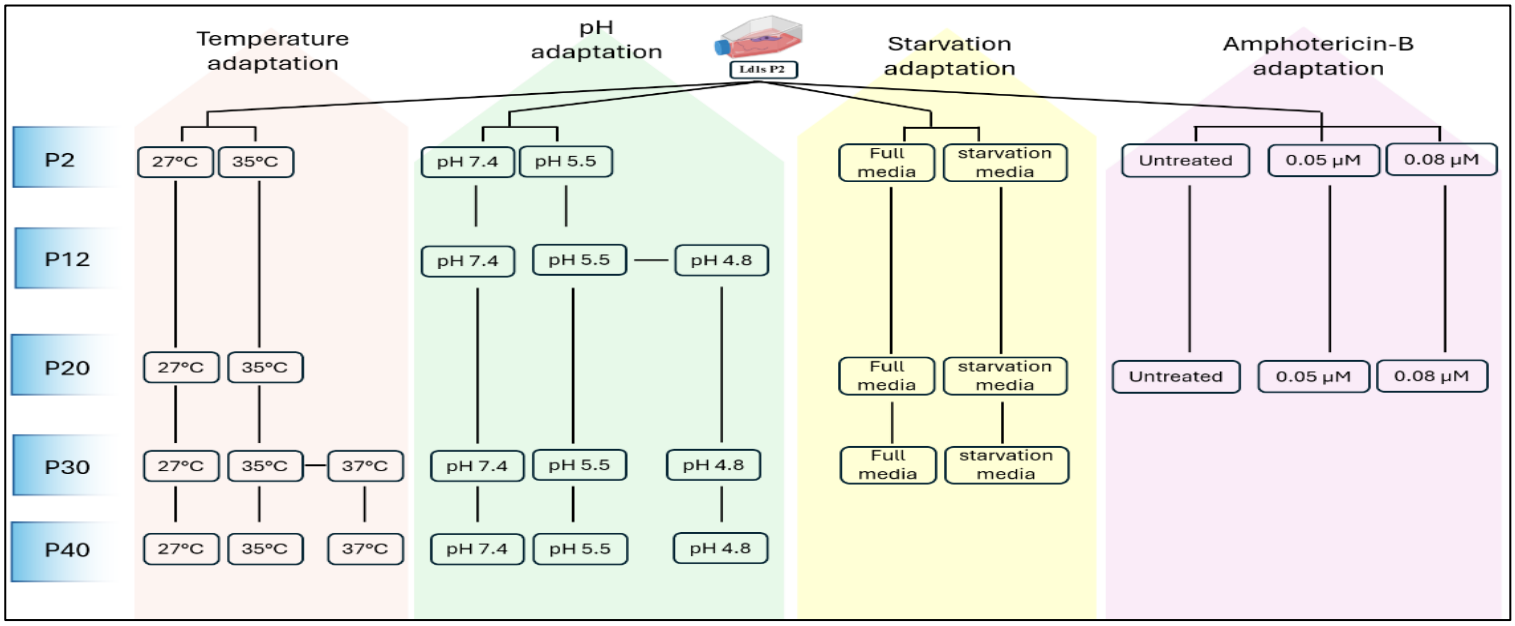


Figure 1B: Laboratory evolution experiment design. *Leishmania donovani* (Ld1S strain) was serially passaged under defined environmental and drug stresses to quantify transcriptional and genomic changes during short- and long-term adaptation. Four selection regimes were used: temperature (27°C control and elevated temperature; with gradual escalation to higher temperature where applicable), pH (pH 7.4 control and acidic pH; with a stepwise decrease to lower pH where applicable), nutrient limitation (full medium vs. starvation medium), and amphotericin B (untreated vs. drug-treated cultures at two concentrations). Selection was initiated at passage 2 (P2), and samples were collected at multiple passages (e.g., P20, P30, P40) for RNA-seq and DNA-seq. Most conditions were profiled with two biological replicates, except the passage 2 and starvation control, which was available as a single replicate.

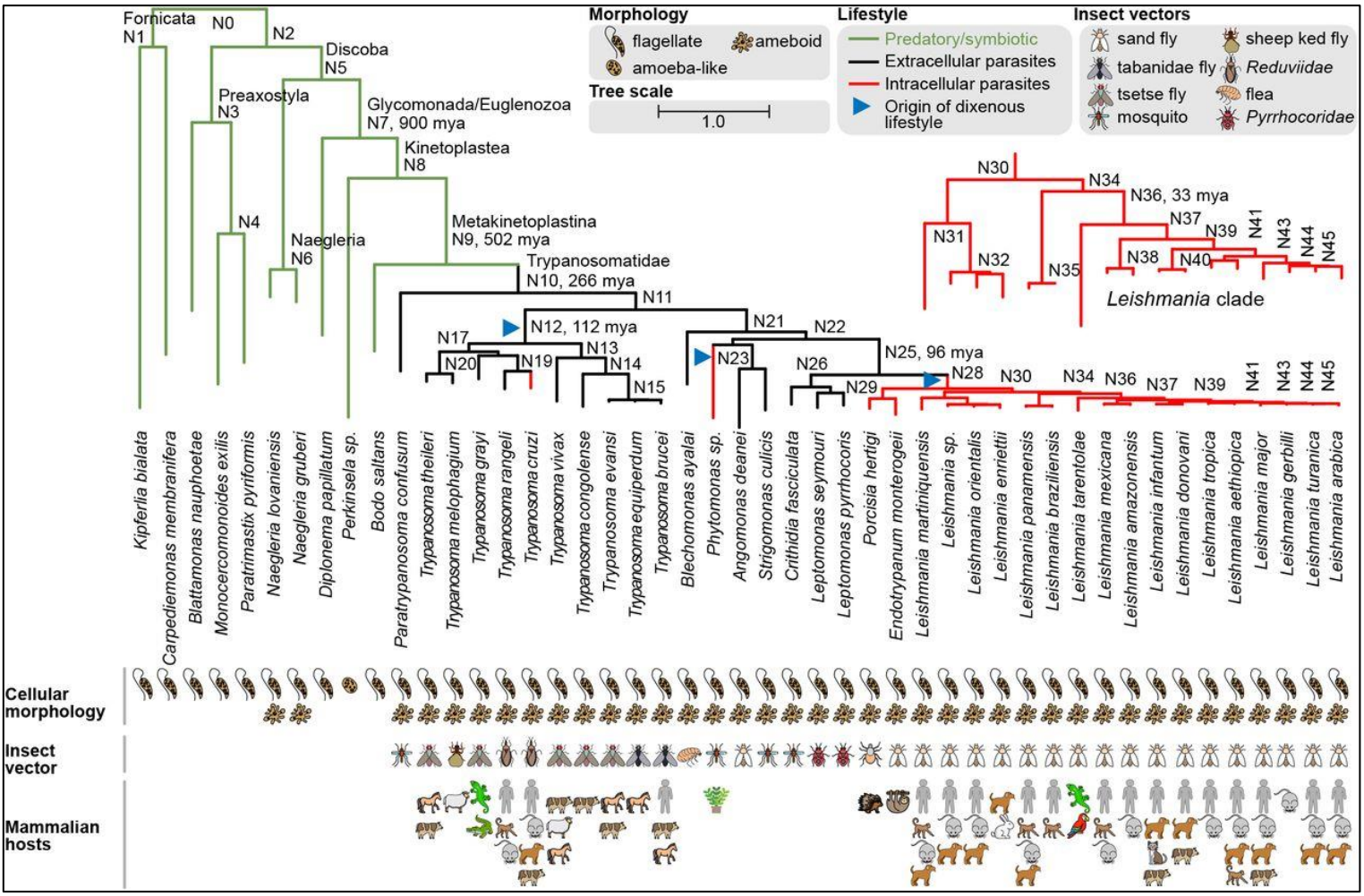


Figure 1C: Evolutionary origin of trypanosomatids from non-parasitic ancestors. Species tree depicting the evolutionary relationships among 47 protists with fully sequenced genomes, including 27 trypanosomatids and 10 predatory or symbiotic taxa. The tree was inferred using maximum-likelihood analysis of a concatenated alignment of 110 single-copy protein families shared across these genomes. Branch lengths represent substitutions per 100 amino acids, and all internal nodes have ≥80% bootstrap support. Where applicable, NCBI Taxonomy-annotated clade names are shown alongside internal node numbers. When available, divergence-time estimates from the TimeTree database are indicated for internal nodes (mya, million years ago). For each organism, known cellular morphologies, insect vectors, and mammalian or plant hosts are illustrated; for monoxenous parasites, only insect vectors are shown. Branches corresponding to intracellular and extracellular parasites are shown in red and black, respectively, whereas non-parasitic protists (predatory or symbiotic) are shown in green. Blue triangles denote parsimoniously inferred origins of the dixenous lifestyle. For clarity, an enlarged view of the *Leishmania* clade is provided.

Goals

Gene expression regulation in *Leishmania* is primarily characterized by chromosome copy number variation and post-transcriptional regulation at the level of individual genes. However, it remains unclear whether the basic structural units of polycistronic transcription represent an additional regulatory layer operating at an intermediate genomic scale. The main goal of this thesis is to investigate whether polycistronic transcription units (PTUs) constitute such an intermediate level of gene expression regulation and to evaluate the contribution of translation efficiency to adaptive expression responses.

To address this goal, the specific objectives of this study were:

1. **To determine whether PTUs exhibit intrinsic structural and functional properties** by analyzing positional gradients of sequence composition, transcriptional output, RNA polymerase II occupancy, mRNA stability, and translation-related features along PTUs.
2. **To test whether PTUs act as independent regulatory units**
   by comparing transcriptional responses of individual PTUs to those of neighboring PTUs and to chromosome-wide expression patterns under diverse environmental and evolutionary conditions.
3. **To assess and compare translation efficiency of genes with respect to the tRNA pool**

   To evaluate similarities and differences among related trypanosomatid species by examining components of translation efficiency, including codon usage, tRNA gene content, and the tRNA adaptation index (tAI), and to assess their relationship to stress-induced transcriptional responses and evolutionary trends.

Together, these objectives aim to refine the current understanding of gene expression regulation in *Leishmania* by integrating transcriptional responses with short-term and long-term evolutionary changes and by evaluating the contribution of translation efficiency as an additional regulatory layer.

Methods

RNA-seq data sets

RNA sequences were obtained from our collaborators. RNA sequencing data generated from the laboratory evolution experiment was first filtered to include only genes present across all datasets analyzed, resulting in a final set of 8,664 genes. With the exception of passage 2 and starvation control samples, all conditions were represented by two biological replicates. Replicates were averaged prior to downstream analyses. Two additional datasets representing developmental and growth phases were obtained from published data[25]. Some datasets were provided in RPKM and were converted to TPM. For each experiment, a threshold was defined based on a peak on the left tail of the TPM distribution, corresponding to the aggregation of undetected transcripts. For each condition, replicate expression measurements were averaged. If only a single replicate column was available for a condition, that column was used directly. This produced one expression vector per condition for downstream analyses. Fold change comparisons were calculated as log2 fold changes.

tRNA Adaptation Index (tAI) analyses

Translation efficiency was quantified using the tRNA adaptation index (tAI), which integrates codon usage with the availability of cognate and wobble-decoding tRNAs. All tAI calculations were performed using custom Python scripts implementing the original tAI framework[26] with explicit incorporation of wobble pairing penalties. Absolute and relative codon adaptiveness

Absolute adaptiveness values (W) were computed for each of the 61 sense codons by summing contributions from all tRNAs capable of decoding the codon, accounting for wobble interactions. For a given codon *c*, W was calculated as:

$$W_c = \sum_i (1 - s_i) \cdot \text{tGCN}_i$$

where *i* indexes tRNAs recognizing codon *c*, s_i is the wobble penalty (from an externally provided wobble interaction table), and tGCN_i is the copy number (or condition-specific abundance) of tRNA *i*. Relative adaptiveness values (*w*) were then derived by normalizing each *W* by the maximum *W*

observed across all codons. Codons with *W = 0* were assigned the geometric mean of nonzero *W* values, preventing undefined values while maintaining relative scaling. Coding sequences were parsed from annotated CDS FASTA files. Only CDSs beginning with a canonical start codon (ATG) and with length ≥ 7 nucleotides were considered. Each CDS was split into codons (excluding stop codons), and gene-level tAI was calculated as the geometric mean of relative adaptiveness values across all codons in the sequence:

$$\text{tAI}_{gene} = \exp\left(\frac{1}{n}\sum_{k=1}^{n}\log\ (w_k)\right)$$

where w_k is the relative adaptiveness of codon *k*. Genes lacking valid codons after filtering were assigned missing values.

For cross-species analyses, a single tAI value per gene was computed using genome-wide tRNA pools. For *L. donovani* laboratory evolution data, tAI was computed separately for each condition using condition-specific tRNA pools, yielding a gene-by-condition tAI matrix.

Codon entropy and genome quality assessment

To assess the reliability of tAI estimates and underlying genome assemblies, codon usage entropy was computed for each gene as a Shannon entropy over codon frequencies, using pseudocount smoothing across all 61 sense codons:

$$H_{gene} = -\sum_{i=1}^{61} p_i \log_2(p_i)$$

where $H_{\text{gene}}$is the Shannon entropy of codon usage for a given gene (computed over the 61 sense codons), and $p_i$is the pseudocount-smoothed frequency of codon $i$in that gene's coding sequence. Codon entropy captures the uniformity of codon usage and is expected to be inversely related to translation optimization. For each species, gene-level codon entropy values were projected against corresponding tAI values. A negative correlation between codon entropy and tAI was used as an internal quality-control criterion, under the expectation that highly optimized genes (high tAI) exhibit lower codon entropy. Species or datasets deviating strongly from this expected relationship were flagged as potentially affected by incomplete genome assemblies or inaccurate tRNA annotations.

Results

Part 1: Polycistronic transcription regulation in *Leishmania*

Global Transcriptomic and Chromosomal Ploidy Changes during Laboratory Evolution

To assess global similarities and differences among samples, principal component analysis (PCA) was performed that allowed embedding on the samples onto a 2D plot that together explain close to 60% of the original variance in the data (**Figure 2A**). The PCA revealed two dominant sources of variation. First, samples segregated primarily according to passage number, forming diagonally separated clusters, indicating that long-term culturing and passage progression exert a substantial effect on global gene expression profiles. Second, within passages, samples exhibited condition-specific separation.

Amphotericin B–treated samples, which were available only in passage 20, showed relatively minor separation from their untreated controls. In contrast, high-temperature samples displayed clear separation from both untreated controls and other stress conditions. In passage 40, which included two elevated temperature treatments, the higher temperature treatment exhibited a smaller transcriptional distance from the control than the moderately elevated temperature.

Low pH conditions showed no separation from controls in passage 30, but were clearly segregated in passage 40; Notably, in passage 30, low-pH–treated samples largely overlapped with untreated controls, while high-temperature–treated samples were clearly distinct. In passage 40, both low-pH–treated samples were separated from untreated controls, but high-temperature–treated samples again showed the largest transcriptional divergence.

Chromosome Copy Number Variation is condition related

Analysis of DNA sequencing data revealed widespread mosaic aneuploidy across samples, consistent with previous observations in *Leishmania*[11]. Classification of chromosome copy numbers by experimental condition indicated that 19 chromosomes remained predominantly euploid across conditions, while the other 17 chromosomes displayed variable and condition-dependent aneuploidy states (**Figure 2B, 2C** respectively**)**. Somy scores (chromosome copy

numbers) were calculated using a published pipeline [27] . Samples were assigned to groups based on their condition type: (i) temperature group: conditions treated at 35°C or 37°C; (ii) pH group: conditions treated at pH 5.5 or 4.8; and (iii) control group: untreated conditions (27°C or pH 7.4). Passages (P30 and P40) were pooled.

Notably, certain chromosomes exhibited pronounced condition-specific copy number changes. For example, chromosome 23 showed a ploidy of approximately 2.75 (consistent with ¼ of population being euploid with 2 copies of the chromosome per cell, and ¾ being trisomic) in untreated samples, increased to approximately 3.5 copies of the chromosome per cell (consistent with, and potentially indicative of an equal representation of trisomic and tetrasomic sub-populations) under low-pH conditions, and decreased to approximately 2 under high-temperature stress. Conversely, Chromosome 16 ploidy is approximately 2 in untreated samples and under low-pH conditions, but under high-temperature stress increased above 3. These observations highlight the dynamic nature of chromosomal dosage as a potential regulatory mechanism during adaptation to environmental stress.

Differential Expression Dynamics Across Conditions and Passages

To characterize transcriptional responses to stress, differential expression analyses were performed using two complementary comparison strategies (**Figure 2D**). First, within each passage, treated samples were compared to their corresponding untreated controls. Second, for conditions present in multiple passages, expression profiles were compared across passages to assess temporal dynamics under sustained condition.

Across most comparisons, log2 fold-change values were concentrated within the range of −2 to 2 (corresponding to 0.25- to 4-fold changes), with distributions centered near zero (**Figure 2E**). This pattern was observed consistently across conditions and passages, indicating broadly similar magnitudes of mRNA level responses. A notable exception was the comparison between starvation samples from passage 20 and passage 30, which exhibited a skewed distribution characterized by widespread upregulation.

As expected, individual comparisons revealed condition-specific sets of upregulated and downregulated genes. Given the recurrence of biologically similar stress conditions across passages, we next examined whether transcriptional responses were reproducible over time. Specifically, we asked whether genes responding to a given stress in one passage exhibited a consistent direction and magnitude of response in subsequent passages (**Figure 2F**). While the major observation is that genes maintained similar trends across passages, some genes exhibited opposing responses. For example, two genes (Ld1S_040798700 and Ld1S_040798800, both of which lack annotation) that were downregulated under low-pH conditions relative to untreated controls in passage 30 were upregulated under the same condition in passage 40, and subset of genes that under high-temperature some genes haven't change in passage 30 but changed 0.5 or 2 fold in passage 40 (e.g. Ld1S_340615800 that contain the domain of unknown function (DUF1935), and Ld1S_330610800 which lack annotation). These observations suggest that transcriptional responses to stress are not static but can evolve over prolonged exposure, potentially reflecting adaptive remodeling of stress response mechanisms.

These descriptive analyses do not resolve whether expression changes arise from chromosome-wide dosage effects, gene-specific regulation, or intermediate regulatory units. In particular this prompted a focused investigation of transcriptional behavior at the level of polycistronic transcription units.

Figure 2

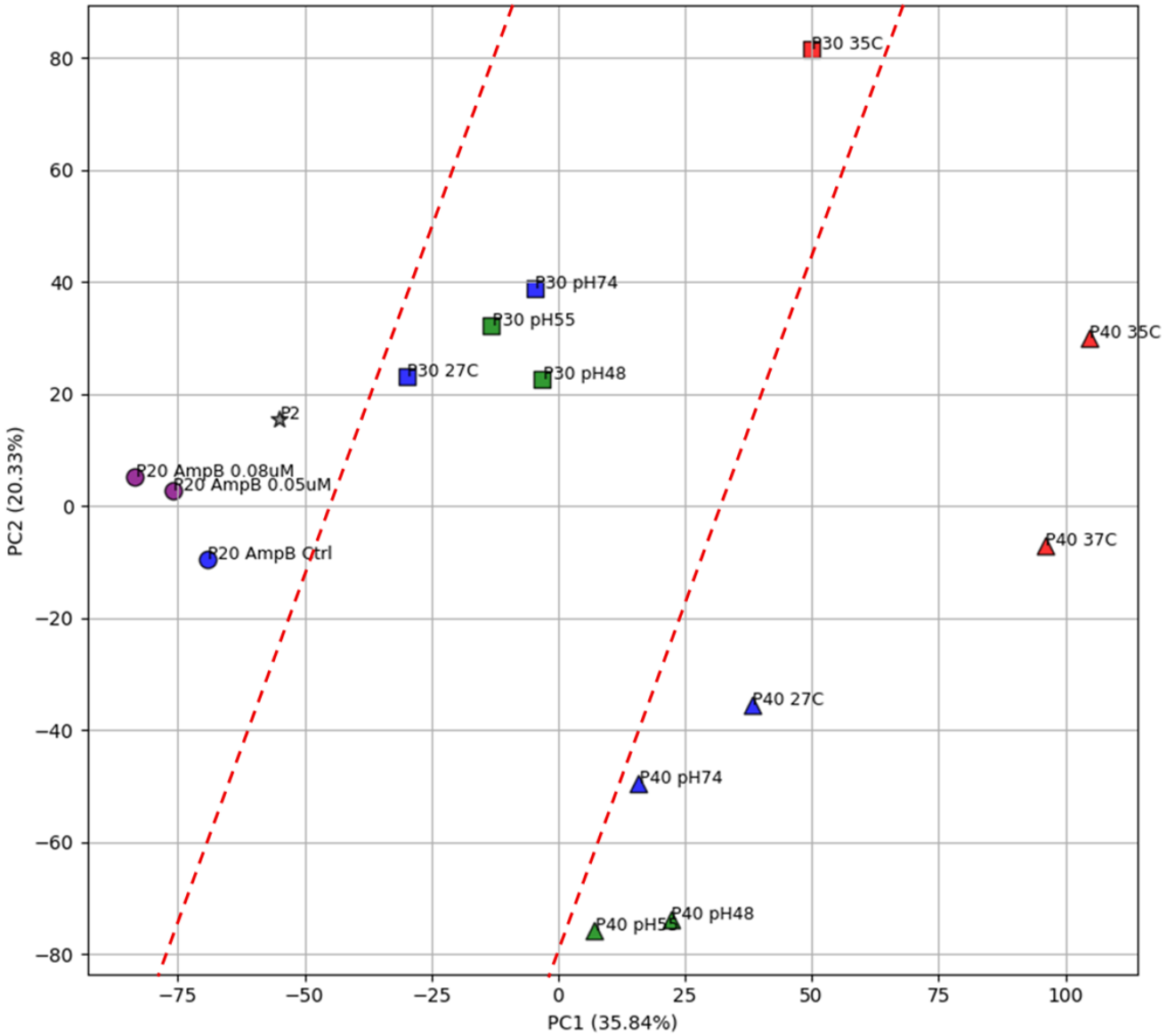


Figure 2A: PCA of the laboratory evolution samples.Red dashed lines indicate the separation between passages. Conditions are distinguished by color: red, high temperature (35°C and 37°C); green, low pH (5.5 and 4.8); purple, amphotericin B (0.05 µM and 0.08 µM); and blue, untreated controls. Replicate measurements were averaged per gene to generate one expression profile per condition. If only a single column existed for a condition, that column was used directly. PCA was performed on the resulting condition-by-gene matrix (conditions as observations, genes as features). Prior to PCA, each gene feature was standardized across conditions to zero mean and unit variance using `StandardScaler` (scikit-learn), ensuring that highly expressed genes did not dominate the decomposition due to scale alone. PCA was then computed using scikit-learn's PCA implementation. To reduce the influence of specific subsets of conditions, PCA was repeated on filtered sets of conditions by excluding predefined groups (e.g., excluding developmental changes, amastigotes, promastigotes, and starvation), while retaining the same preprocessing and PCA procedure.

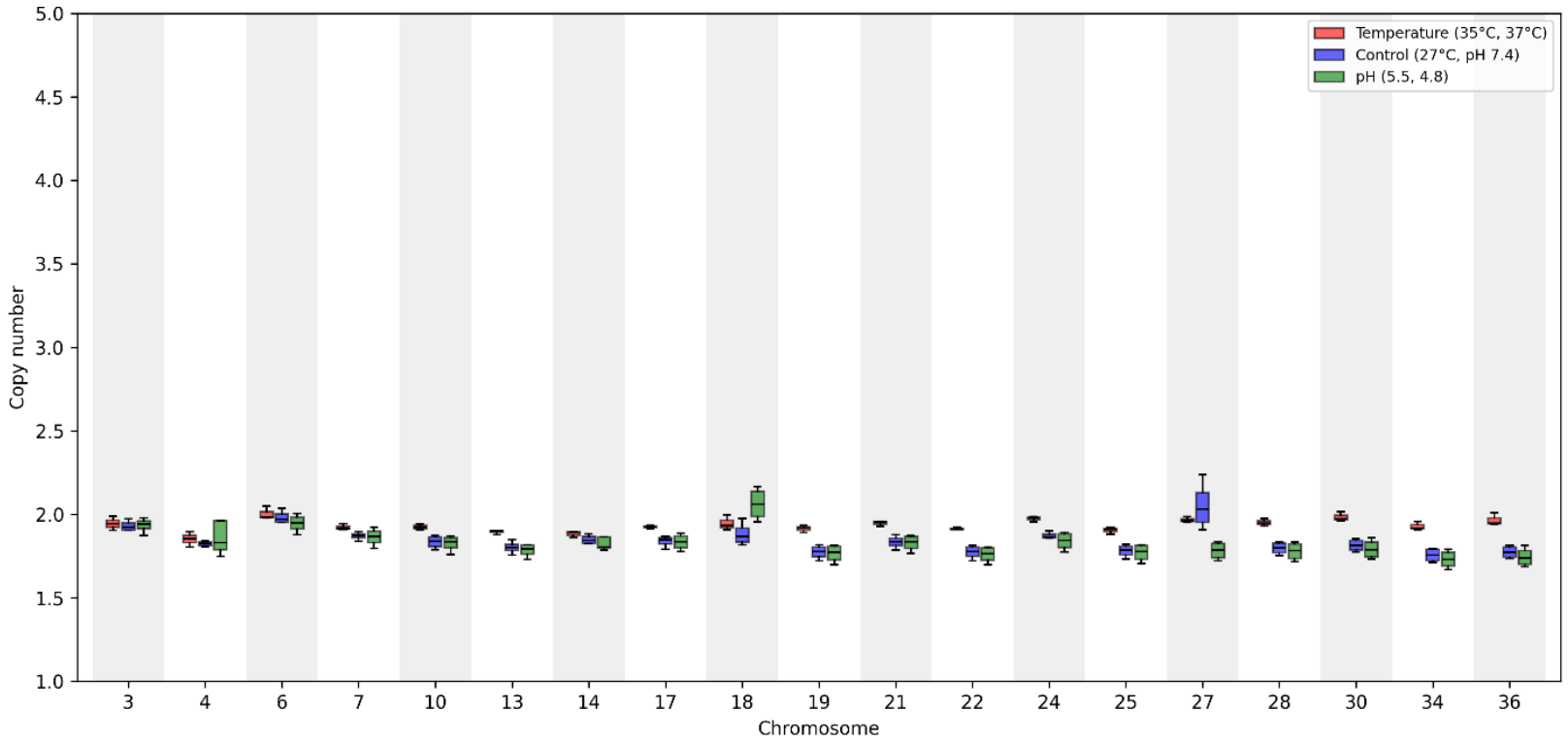


Figure 2B: Chromosome copy number (CCN). This subfigure shows chromosomes with only minor deviations from disomy (1.65 < CCN < 2.35). Sample color coding is consistent with the previous figures.

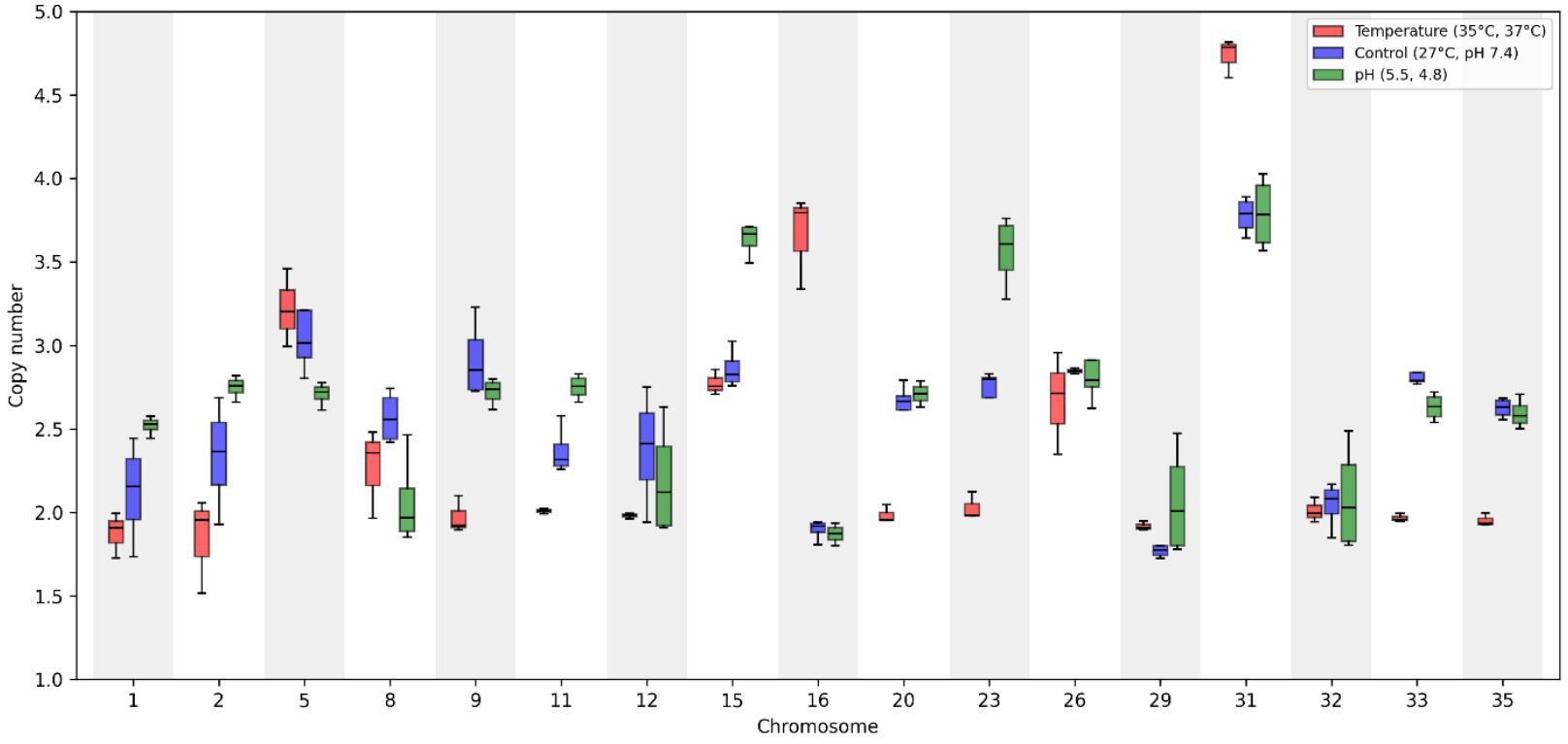


Figure 2C: Chromosome copy number (CCN). This subfigure shows chromosomes with substantial deviations from disomy (CCN < 1.65 or CCN > 2.35). The somy pattern is both chromosome-dependent and condition-dependent.

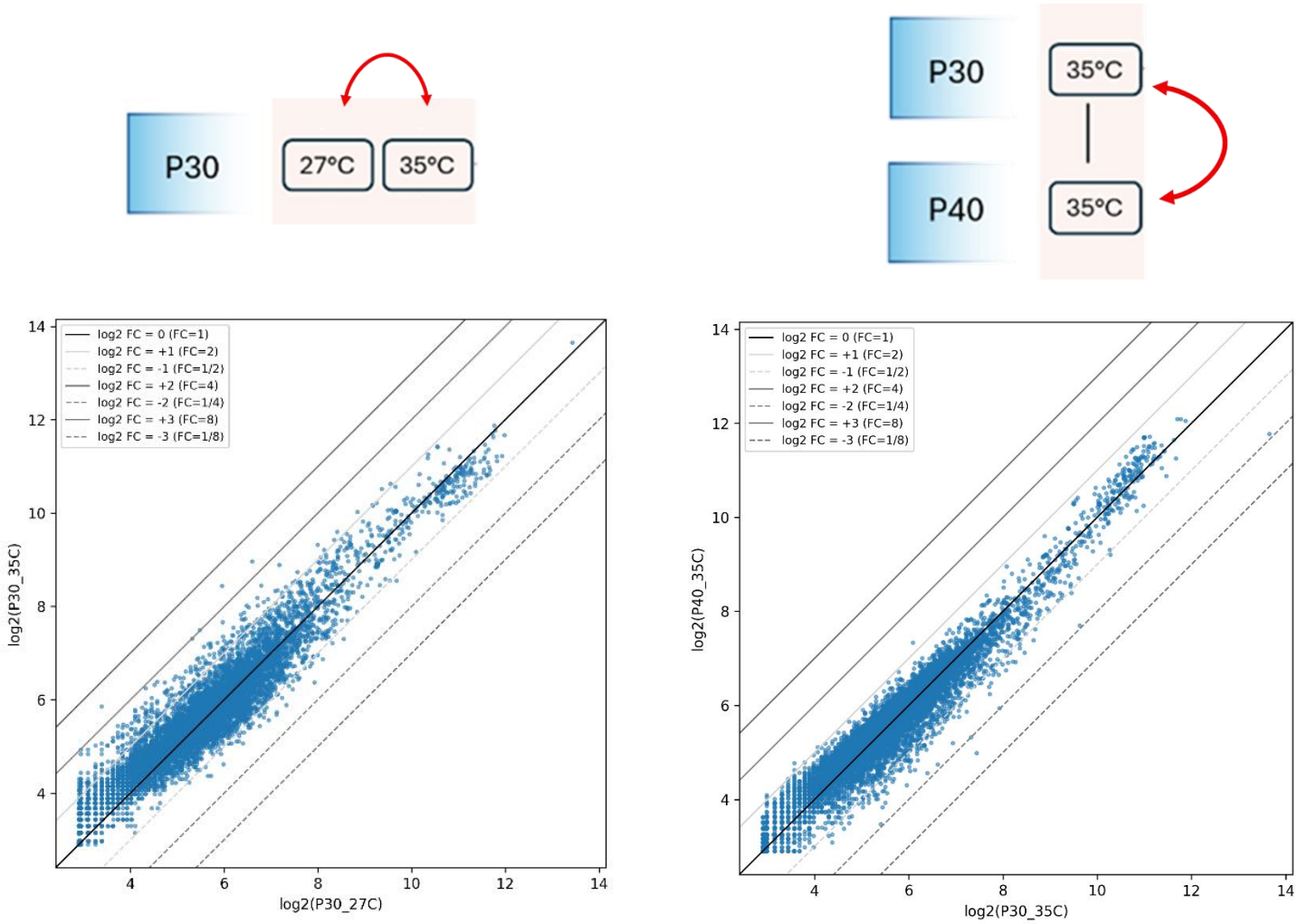


Figure 2D: Differential expression comparisons. Two comparison types were performed. First, within-passage comparisons contrasted each stressed sample with its matched control from the same passage to quantify the transcriptional response to that stress (left). Second, between-passage comparisons contrasted samples from different passages under the same condition to quantify evolution-associated expression changes independent of the specific stress (right). Plotting gene expression values (TPM) on the two axes allows fold changes to be visualized relative to the diagonal lines.

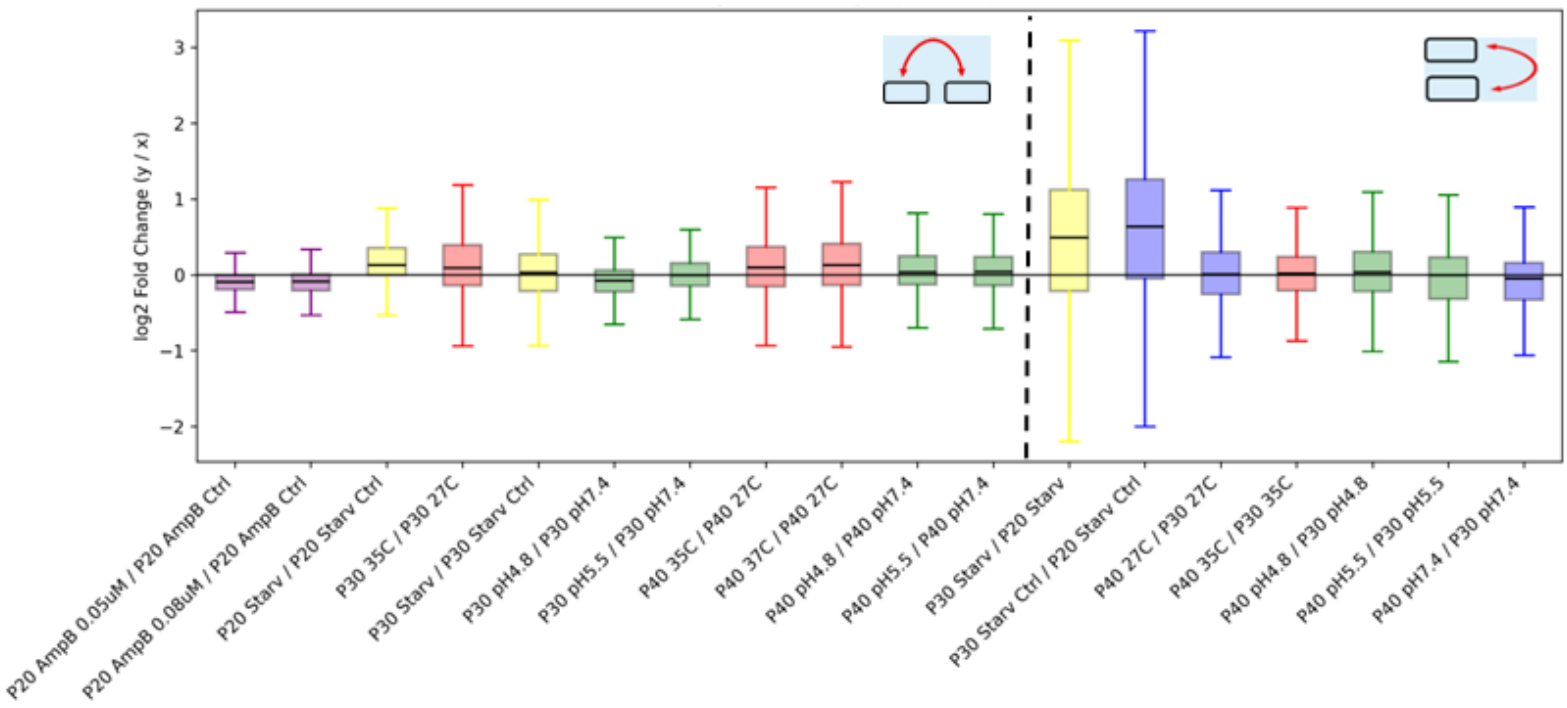
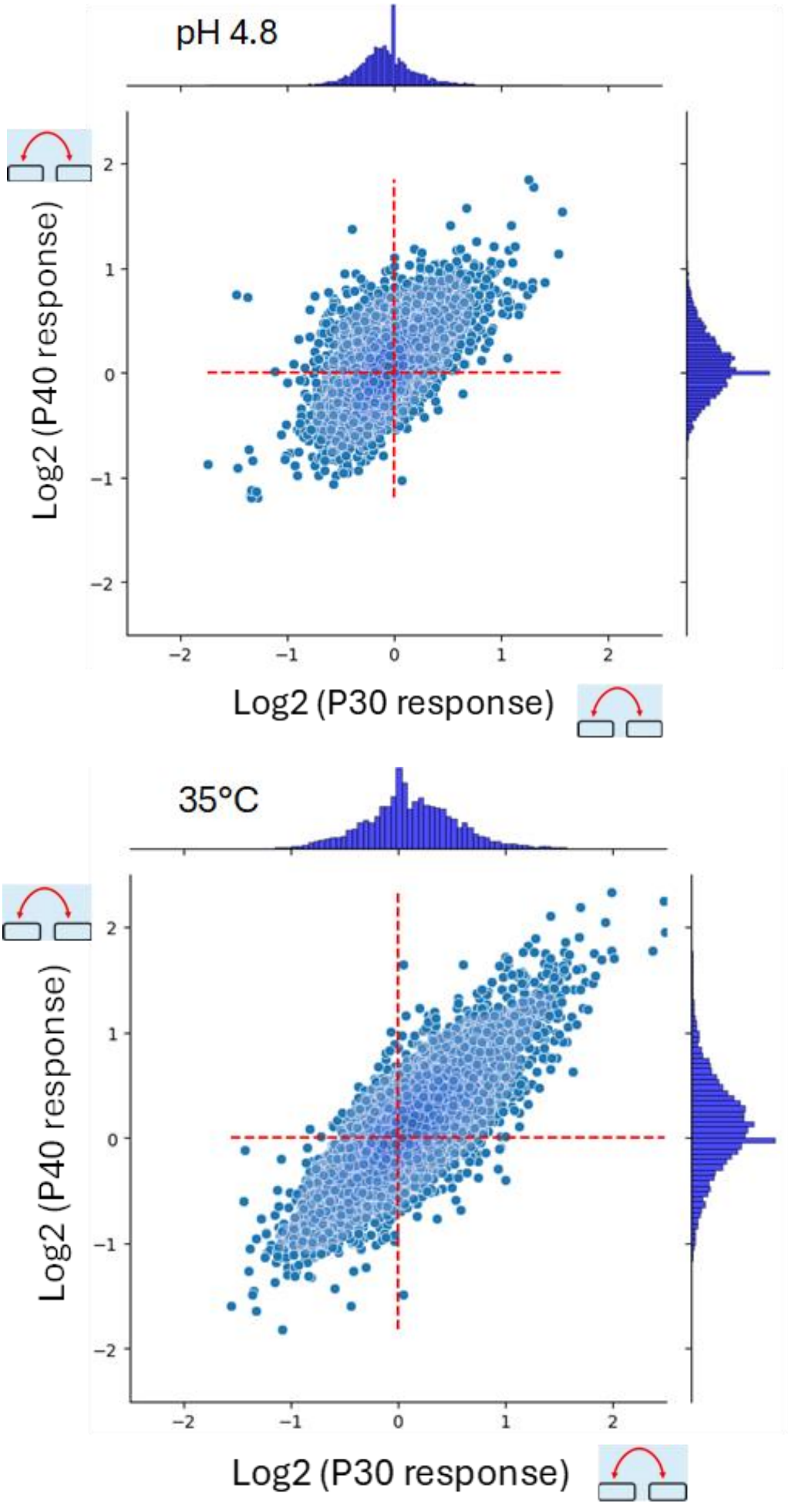


Figure 2E: Span of fold changes. Across most comparisons, fold changes were modest, typically ranging from log2 −1 to 1 (0.5-fold to 2-fold). In the within-passage comparisons, the temperature-adaptation contrasts tended to show a slightly broader range than the other conditions. In the between-passage comparisons, the starvation contrasts and the starvation-control contrasts showed a larger and distinct range (log2 −2 to 3; 0.25-fold to 8-fold).

Figure 2F: Consistency of changes. These panels provide representative examples of how adaptive transcriptional responses evolve over time by comparing passage 30 versus passage 40; additional comparisons for the same stressors across other passages are shown elsewhere. For the response to pH 4.8 (top), most genes respond similarly in both passages (Pearson r = 0.59), although a small subset shows an opposite response. For example, in the upper-left quadrant, two genes are downregulated in passage 30 (log2 fold change = −1.47 and −1.36) but upregulated in passage 40 (log2 fold change = 0.74 and 0.72). Both genes lack functional annotation. For the response to 35°C (bottom), most genes also follow a consistent trend between passages (Pearson r = 0.81), with a small number deviating from the overall relationship. Notably, some genes are essentially unresponsive in one passage (log2 fold change ~ 0; near the red dashed axes) but show a clear response in the other passage, for example (P30, P40) log2 fold changes of (0.05, 1.64) and (0.05, −1.48) for two genes lacking annotation, and (−1.43, 0.01) for a gene annotated as peptidyl-dipeptidase.



PTU Structure exhibit Consistent Positional effect on transcription

To investigate whether *Leishmania* can regulate transcription at the level of polycistronic transcription units (PTUs), in contrast to prevailing dogma in this species field, we first defined PTU boundaries in a consistent and biologically informed manner. Strand-switch regions, where the orientation of consecutive genes changes, necessarily demarcate transcriptional unit boundaries and were therefore used to identify divergent and convergent PTUs. However, for tandem PTUs arranged head-to-tail, gene orientation alone is insufficient to define boundaries.

To resolve tandem PTU boundaries, we incorporated transcription start site (TSS) annotations derived from ChIP-seq experiments in *Leishmania tarentolae* and *Leishmania infantum*[28,29]. These annotations are based on enrichment of histone variants and post-translational modifications associated with transcription initiation. Given the high conservation of transcriptional architecture across *Leishmania* species, we assumed conservation of TSS positions and used these data to infer PTU boundaries in the analyzed genomes.

To characterize PTU-intrinsic structure, we constructed meta-PTU profiles by aligning genes according to their relative position within each PTU. For each feature, values within individual PTUs were normalized to the PTU mean, and signals were averaged across all PTUs to obtain position-specific meta-profiles.

Across multiple molecular features, a highly consistent positional gradient was observed along PTUs. **GC content** (**Figure 3A, top**) exhibited a characteristic pattern, with relatively low GC content at the very first gene positions, a peak at early PTU positions (approximately the fifth gene), and a gradual decline toward the PTU end.

Expression levels (**Figure 3A, middle**) Meta-PTU profiles of transcript abundance revealed higher expression at PTU beginnings, with expression decreasing along the PTU. This pattern was observed in both *Leishmania major* and *Trypanosoma brucei*, indicating conservation across trypanosomatids. Similar trend was observed also in nascent RNA data from *L. major.* It is possible that some of these trends reflect RNA polymerase drop-off, where a fraction of elongating polymerases abort transcription before reaching the end of the PTU. Such premature termination would enrich signal near PTU beginnings and contribute to the apparent decline in transcript abundance toward downstream regions.

mRNA half-life (**Figure 3B, right**) displayed a similar gradient, with higher transcript stability near PTU starts and progressively lower stability toward downstream positions.

RNA polymerase II occupancy (**Figure 3B, left**) also displayed a PTU-embedded positional pattern. In *L. major*, polymerase occupancy derived from nascent RNA sequencing (PRO-seq)[30] showed a gradual decrease from PTU starts toward PTU ends. A comparable gradient was observed in *T. brucei* based on RNA polymerase II ChIP-seq data. Together, these results demonstrate that transcriptional output and polymerase distribution are coordinated with PTU architecture.

Translation efficiency, quantified using the tRNA adaptation index (tAI), also showed elevated values at the beginning of PTUs followed by a gradual decline with increasing distance from the transcription start site. These gradients indicate that sequence composition, transcript stability, and translational potential are all systematically associated with PTU structure. Consistent with this interpretation, analysis in *Schizosaccharomyces pombe* revealed a genome-wide positive association between codon optimality and mRNA half-life, with the relationship remaining significant when codon optimality was quantified by tAI[31]. Lower tAI values toward downstream PTU regions may therefore promote reduced mRNA stability and contribute to the observed expression-level gradient.

To assess whether PTU-associated positional effects generalize at the transcriptome scale, we compared the relative PTU positions of the 200 highest-expressed and 200 lowest-expressed genes for each condition. Significant differences in PTU position distributions were observed in 20 out of 26 analyzed conditions, with highly expressed genes preferentially localized toward PTU beginnings (**Figure 3C**). This supports the generality of PTU-embedded gradients across diverse environmental conditions.

Finally, we examined whether intra-PTU positioning is associated with gene function. Analysis of enzyme pairs participating in the same metabolic pathways revealed partial conservation of relative positions in PTU relative to their start sites. Similarly, subunits of heteromeric protein complexes tended to occupy comparable positions within PTUs. Although these correlations are moderate, they suggest that PTU organization may contribute to coordinated regulation of functionally related genes (**Figure 3D**). This positional coordination may indicate that genes occupying similar distances from

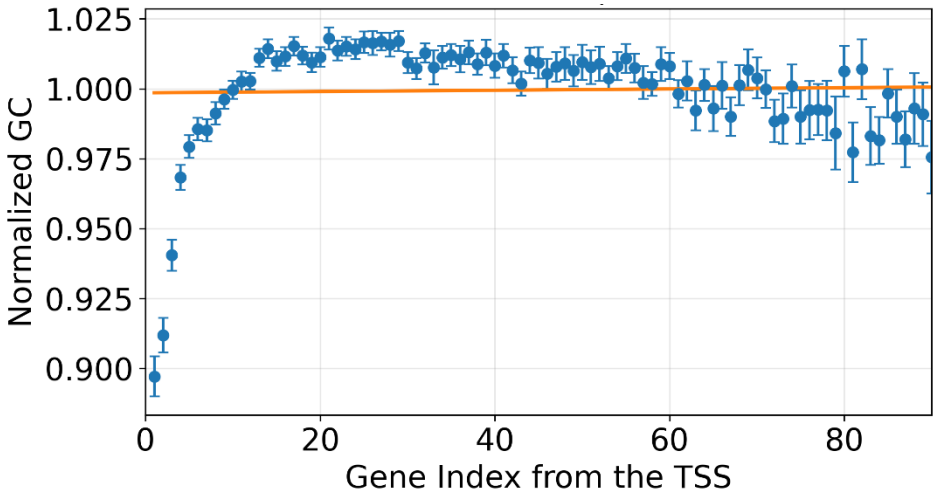
the PTU start site are preferentially co-transcribed, thereby promoting synchronized expression. Such synchronization could facilitate efficient metabolic flux through shared pathways and support balanced production of complex subunits needed for effective assembly.

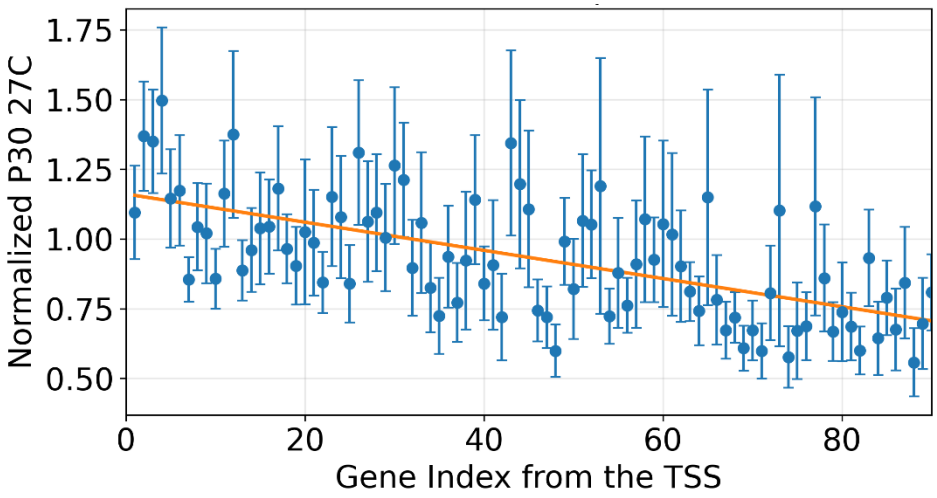
The presence of conserved positional gradients across multiple molecular features indicates that PTUs possess intrinsic structural and regulatory properties that are maintained across species and experimental conditions. However, these analyses alone do not determine whether PTUs act as independent regulatory units in response to environmental stress or whether their behavior merely reflects chromosome-wide dosage effects or gene-specific post-transcriptional regulation. To address this question, we next examined whether transcriptional responses of individual PTUs diverge from those of neighboring PTUs and from the overall response of the chromosome on which they reside. This analysis directly tests the hypothesis that PTUs constitute an intermediate, independently regulated layer of gene expression control.

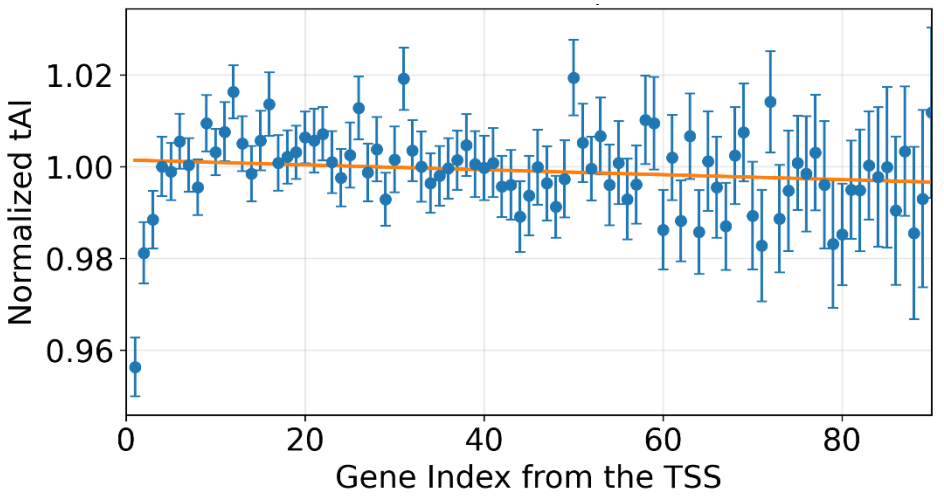
Figure 3

Figure 3A: PTU profiles in L. donovani. For each metric, values were normalized within each PTU by dividing the per-position value by the PTU mean. Dots indicate the mean across PTUs at each position, and error bars indicate the standard error of the mean. Gene index from TSS values were restricted to a fixed window by retaining only genes within positions 1–90. GC content was computed per gene as the fraction of valid DNA bases (A/C/G/T) that were G or C. PTUs <20 genes were filtered.

Top: GC content. At the beginning of the PTU, the first four genes show distinctly low GC content; GC then increases to a peak around the ~15th gene and subsequently declines gradually. The overall Pearson correlation is r = 0.03, but because the pattern is clearly non-linear, this summary statistic is not informative. **Middle**: Transcription rates in a representative sample (passage 30 at 27°C). The overall trend shows a decreasing gradient along the PTU (Pearson r = −0.61, p = 9.42e−11). Across all conditions included in the analysis, Pearson r ranged from −0.61 to −0.27, with a mean of −0.49. **Bottom**: tAI profile. Consistent with the GC pattern, tAI starts relatively low, reaches a peak, and then declines gradually. The overall Pearson correlation is r = −0.152, but as for GC content, the non-linear profile makes this global correlation difficult to interpret.







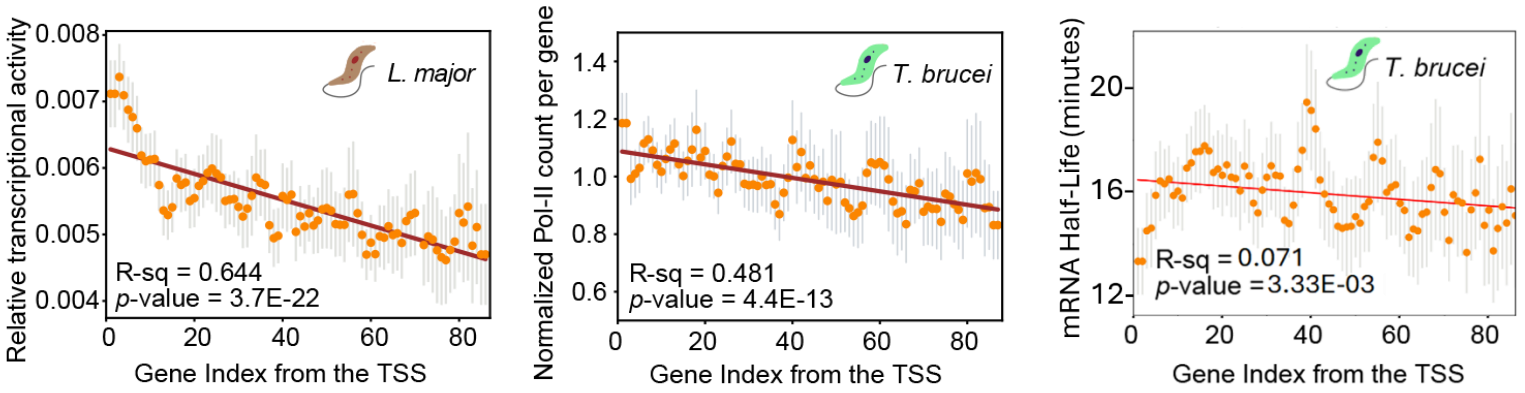


Figure 3B: PTU profiles in *L. major* and *T. brucei*. Values were normalized as in Figure 3A, by dividing the per-position value within each PTU by the PTU mean. **Left:** relative transcriptional activity, derived from a PRO-seq experiment in *L. major*. **Middle:** RNA polymerase II (Pol II) occupancy in *T. brucei*. **Right:** mRNA half-life in *T. brucei*. These panels are from our paper in preparation.

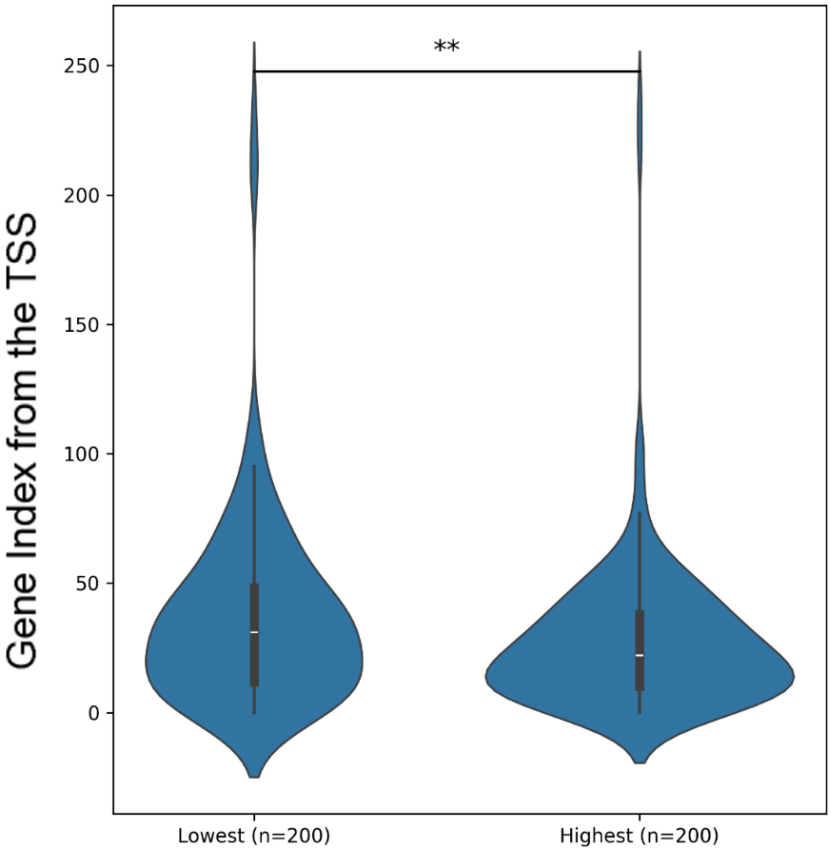


Figure 3C: Genome-scale effect of PTU position on expression. Shown is the distribution of within-PTU positions for the 200 highest expressed genes and the 200 lowest expressed genes in passage 30 at 27°C. The highest expressed genes are significantly enriched near the beginning of PTUs compared with the lowest expressed genes.

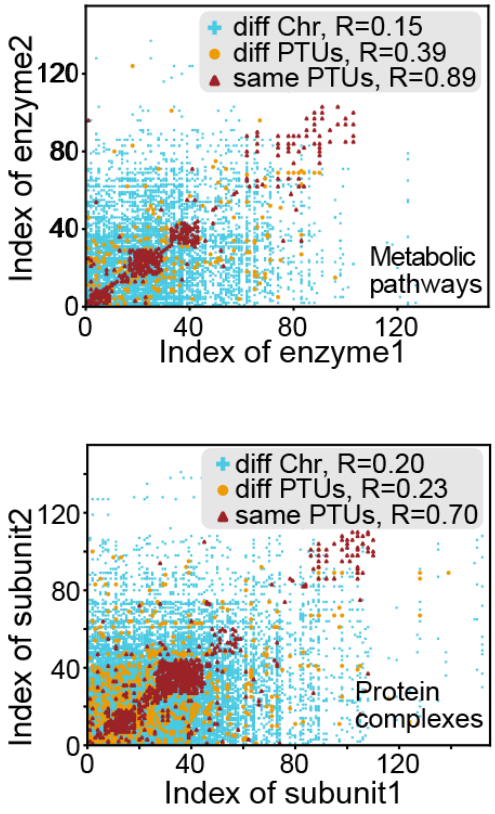


Figure 3D: PTU position correlates within enzymatic pathways and protein complexes. Top: correlation of PTU positions for pairs of enzymes participating in the same metabolic pathway. Bottom: correlation of PTU positions for pairs of subunits belonging to the same protein complex. Color code: red, pairs located within the same PTU; orange, pairs located on the same chromosome but in different PTUs; light blue, pairs located on different chromosomes.

Statistical testing for differential expression reveals PTU independence

After defining polycistronic transcription units (PTUs) based on gene orientation and conserved transcription start sites inferred from histone modification data, and after establishing that PTU structure contributes systematically to expression-related features, we next asked whether PTUs also function as independent regulatory units. Specifically, we tested whether genes belonging to the same PTU exhibit coordinated regulatory responses that differ from those of neighboring PTUs and from the overall response of other genes in their same chromosome.

Current models of gene regulation in *Leishmania* emphasize two dominant scales: chromosome-wide regulation mediated by changes in copy number, and gene-level regulation through post-transcriptional mechanisms. If PTUs represents an additional regulatory layer, we would expect the transcriptional response of a given PTU to be distinct from that of its adjacent PTUs and / or from the chromosome-wide background, even when accounting for these known regulatory mechanisms.

To address this question, we analyzed gene-level log2 fold changes derived from multiple biologically distinct comparisons, including acute stress responses, long-term experimental evolution, and developmental or growth-phase transitions. For each comparison, PTU-level independence was assessed by testing whether the distribution of gene-level fold changes within a PTU differed significantly from its immediate upstream neighbor, its immediate downstream neighbor, and the remainder of the chromosome.

Overall, we observed diverse patterns rather than a sharp binary outcome, with these features changing dynamically across conditions and across PTUs (**Figure 4A**). Some PTUs displayed multiple independences across multiple experimental contexts, whereas others were independent only under specific environmental or evolutionary conditions.

When comparisons were stratified by experimental context, passage contrasts (the same condition measured at different passages; e.g., 27°C at P40 versus P30) yielded a stronger PTU-independence signal than treatment (different conditions within the same passage; e.g., P30 at 35°C versus 27°C). Consistent with this trend, the mean and median percentages of independently regulated PTUs were 12.5% and 10.9% for passages comparisons, compared with 8.4% and 7.8% for treatment comparisons. This pattern suggests that expression changes associated with long-term adaptation across passages may involve PTU-level regulation more prominently than within-

passage treatment contrasts, which more closely reflect immediate physiological responses to stress. It should be noted, however, that the treatment contrasts analyzed here are not strictly “pure” physiological comparisons, because treated and untreated populations were propagated separately and therefore may already differ due to short-term adaptation. A cleaner assessment of purely physiological stress responses would compare a non-adapted control population to a non-adapted population exposed acutely to elevated temperature, prior to sustained passaging and adaptation.

Overall, among the 200 PTUs examined, 134 PTUs (67%) were identified as independently regulated in at least one comparison. Each such case of a PTU compared to its surrounding PTUs, or chromosome wide neighborhood was assessed a p-value (two-sided Mann–Whitney U tests) and corrected for multiple hypotheses testing by FDR. Importantly, these signals remained robust after filtering extreme gene-level outliers using a modified Z-score threshold, indicating that PTU-level independence generally reflects coordinated behavior across multiple genes rather than isolated extreme responders.

Together, these results demonstrate that PTU-level regulation is a prevalent and systematic feature of the *L. donovani* transcriptome. Rather than acting solely as passive structural units, PTUs frequently exhibit transcriptional responses that are distinct from both their chromosomal background and their immediate genomic neighbors, supporting the existence of an intermediate regulatory layer between chromosome-scale and gene-scale regulation.

Cross-Species and Cross-Platform Validation

To assess whether PTU independence is conserved across trypanosomatids, we applied the same analytical framework to RNA-seq datasets from *Trypanosoma brucei* and *Leishmania major*. For *T. brucei*, six publicly available RNA-seq datasets were analyzed. Across all comparisons, approximately 15% of PTUs were detected as independently regulated. This lower fraction may reflect differences in experimental conditions, the limited number of datasets analyzed, or intrinsic differences in genome organization, as *T. brucei* has only 11 chromosomes and more than 500 PTUs.

For *L. major*, PTU independence was evaluated using PRO-seq data, which measures nascent RNA associated with actively transcribing RNA polymerase II. Unlike RNA-seq–based comparisons, this analysis did not rely on differential expression between conditions but instead compared absolute

transcriptional output across PTUs. Consequently, the regulatory signal assessed here reflects transcription initiation and elongation prior to the influence of post-transcriptional processes.

In this context, approximately 73% of PTUs in *L. major* exhibited transcriptional profiles that differed significantly from their surrounding genomic regions (**Figure 4B**). This result is notable for the widespread differences between PTUs suggest that transcriptional activity varies substantially between PTUs, consistent with differential regulation of transcription initiation or polymerase engagement.

Validation of PTU Boundary Specificity

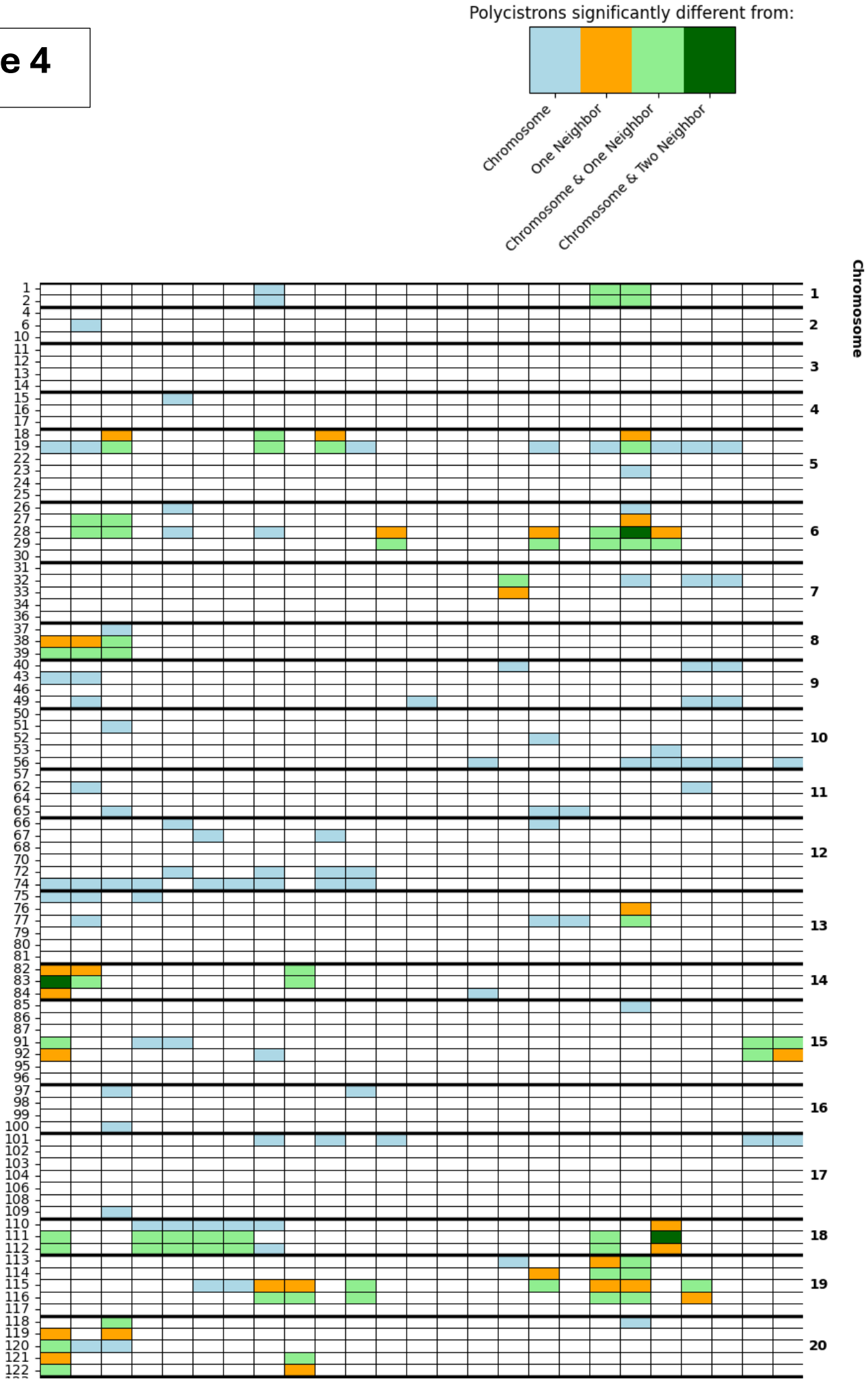
To assess whether the observed PTU-level regulatory independence in Leishmania major could arise merely due to chromosome-wide signal structure or PTU size distributions, as opposed to bona fide PTU –based regulation, we performed two randomization-based negative controls. In the first randomization, gene positions were shuffled but PTU boundaries were preserved. In the second randomization, chromosome structure and gene order were preserved, but PTU boundaries were randomly reassigned, keeping fixed the distribution of the length of the PTUs, measured in gene number per PTU.

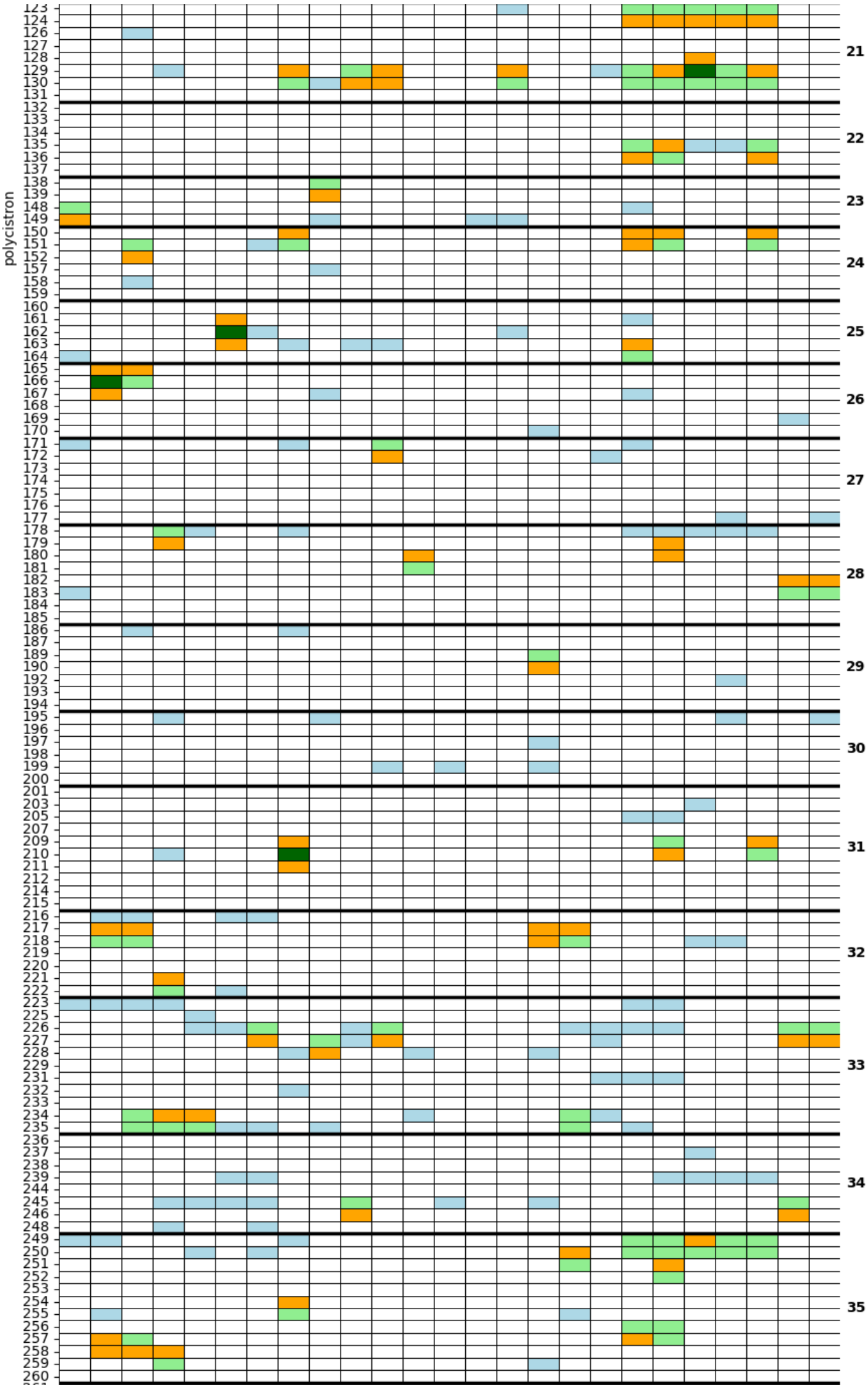
In both randomizations, the proportion of independently regulated PTUs detected in the randomized data was significantly lower than that observed using the true PTU definitions (**Figure 4C**).

This result supports the conclusion that PTU independence arises from genuine biological organization defined by gene orientation and transcription start site architecture, rather than from random clustering effects.

The widespread and condition-specific independence of PTU regulatory responses suggests that PTUs are not merely passive transcriptional units but may be selectively activated under defined environmental conditions. This raises the question of whether PTU-specific regulation is mediated by localized regulatory elements, such as promoters or other cis-acting sequence features associated with transcription initiation. To explore this possibility, we next investigated whether independently regulated PTUs share common sequence motifs or compositional features that could underline coordinated activation, with a particular focus on temperature-induced transcriptional responses.

Figure 4


Polycistrons significantly different from:
Chromosome
One Neighbor
Chromosome & One Neighbor
Chromosome & Two Neighbor
Chromosome
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20


polycistron
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35

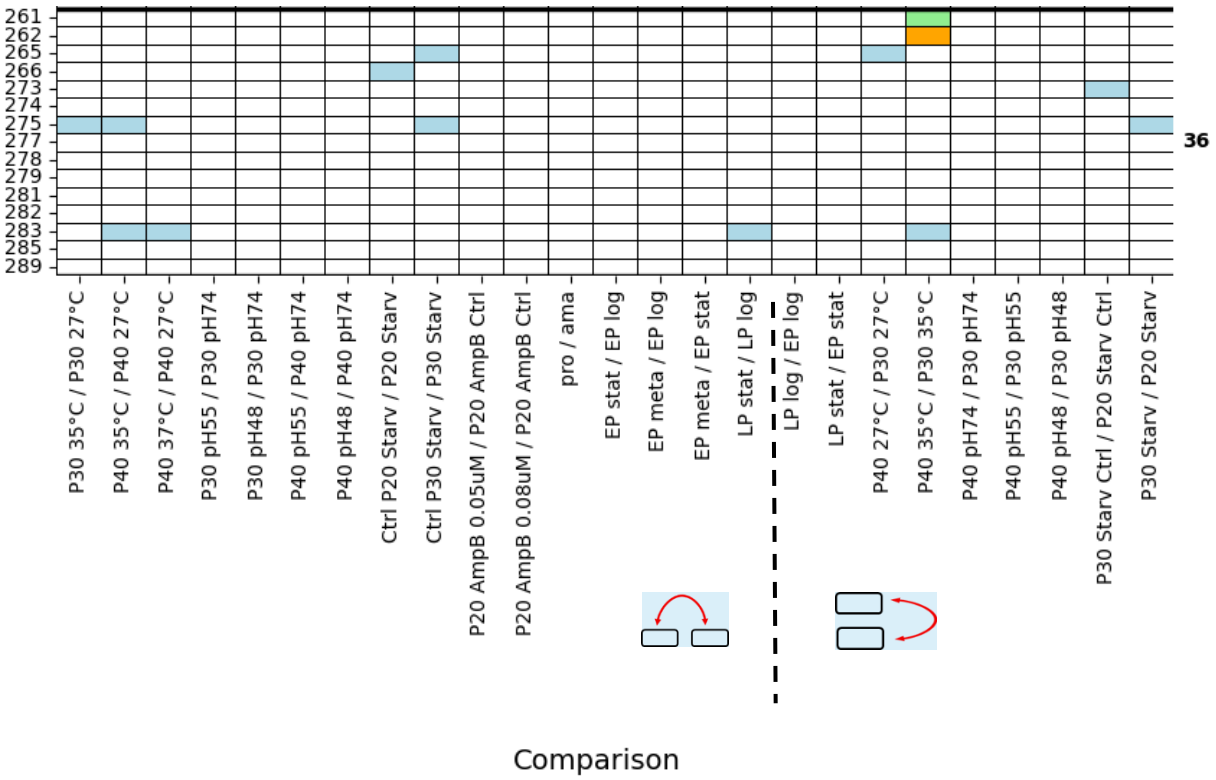


Figure 4A: Matrix of independent PTUs. Within-passage comparisons are shown on the left, and between-passage comparisons on the right. Between-passage comparisons contained a higher fraction of independent PTUs (median 10.9% vs. 7.8% for within-passage comparisons). Chromosome boundaries are indicated by black horizontal lines, and chromosome labels are shown on the right y-axis.

All statistical comparisons were performed using two-sided Mann–Whitney U tests. Tests were conducted only when both compared distributions contained more than two genes. p-values were corrected for multiple testing using the Benjamini–Hochberg false discovery rate (FDR) procedure across all PTUs. PTUs with FDR-adjusted p-values below 0.05 were considered significantly independent. For significant PTUs, the direction of regulation relative to the reference distribution (chromosome background or neighboring PTU) was determined by comparing median log2FC values and classified as Up or Down. To ensure that PTU-level independence was not driven by single extreme genes, all significant comparisons were re-evaluated after removing gene-level outliers using a modified Z-score approach. Genes with an absolute modified Z-score greater than 3.5 were excluded in a direction-aware manner (high outliers for upregulated PTUs, low outliers for downregulated PTUs).

Color code: light blue-PTU log2 fold-change distribution differs from its chromosome; orange-PTU log2 fold-change distribution differs from one neighboring PTU; light green-PTU log2 fold-change distribution differs from its chromosome and one neighboring PTU; dark green-PTU log2 fold-change distribution differs from its chromosome and both neighboring PTUs.

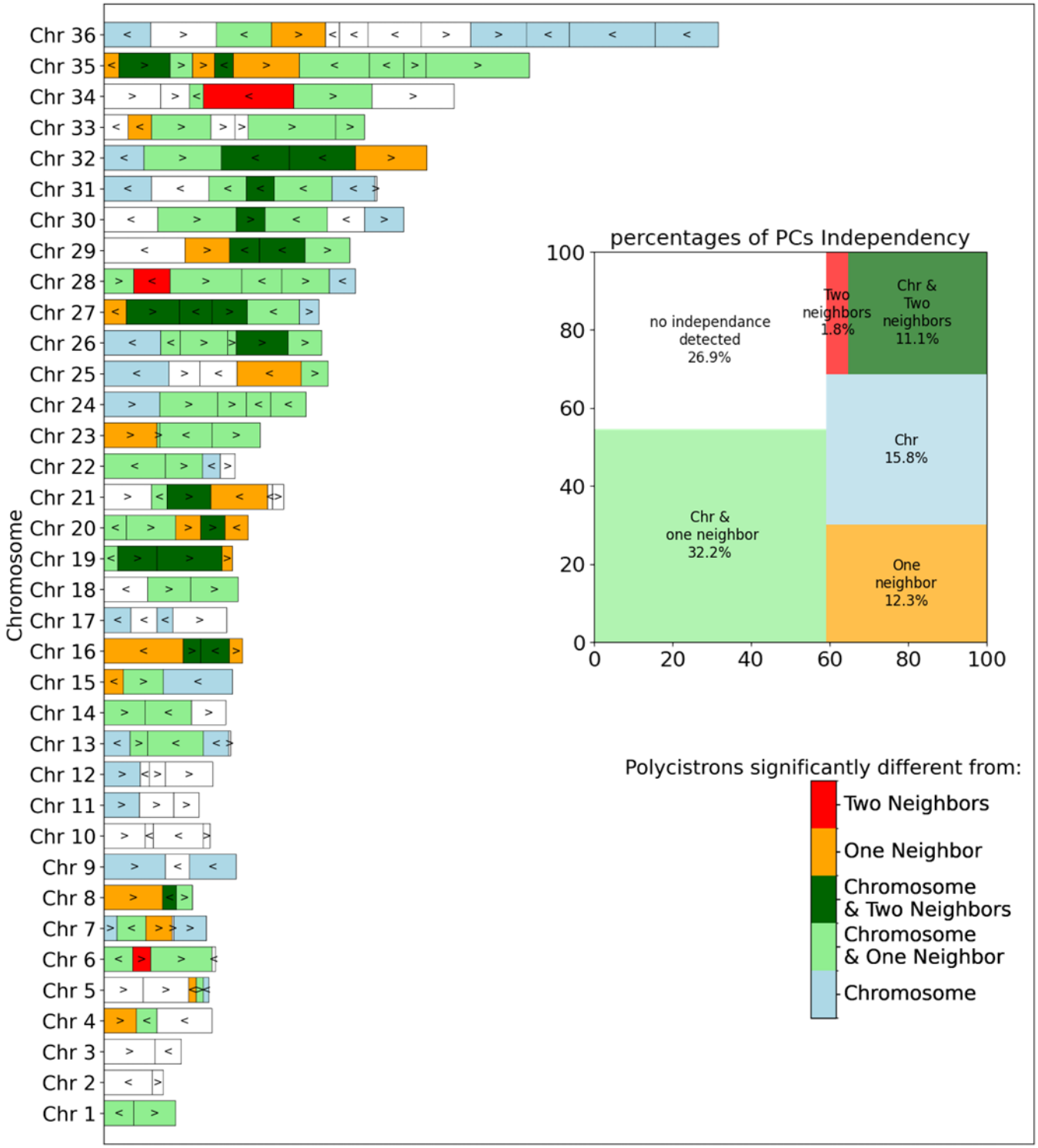


Figure 4B: Independent PTUs in *L. major* PRO-seq. Polycistronic transcription units (PTUs) are shown as rectangles arranged along each chromosome. PTU orientation is indicated by arrows: ">" for the plus strand and "<" for the minus strand. Rectangle width is proportional to the number of genes in the PTU. Colors (see legend) denote the class of independence detected for each PTU, based on significant differences from the chromosome background and/or from one or both neighboring PTUs. The treemap summarizes the proportion of PTUs assigned to each independence class among PTUs included in the analysis.

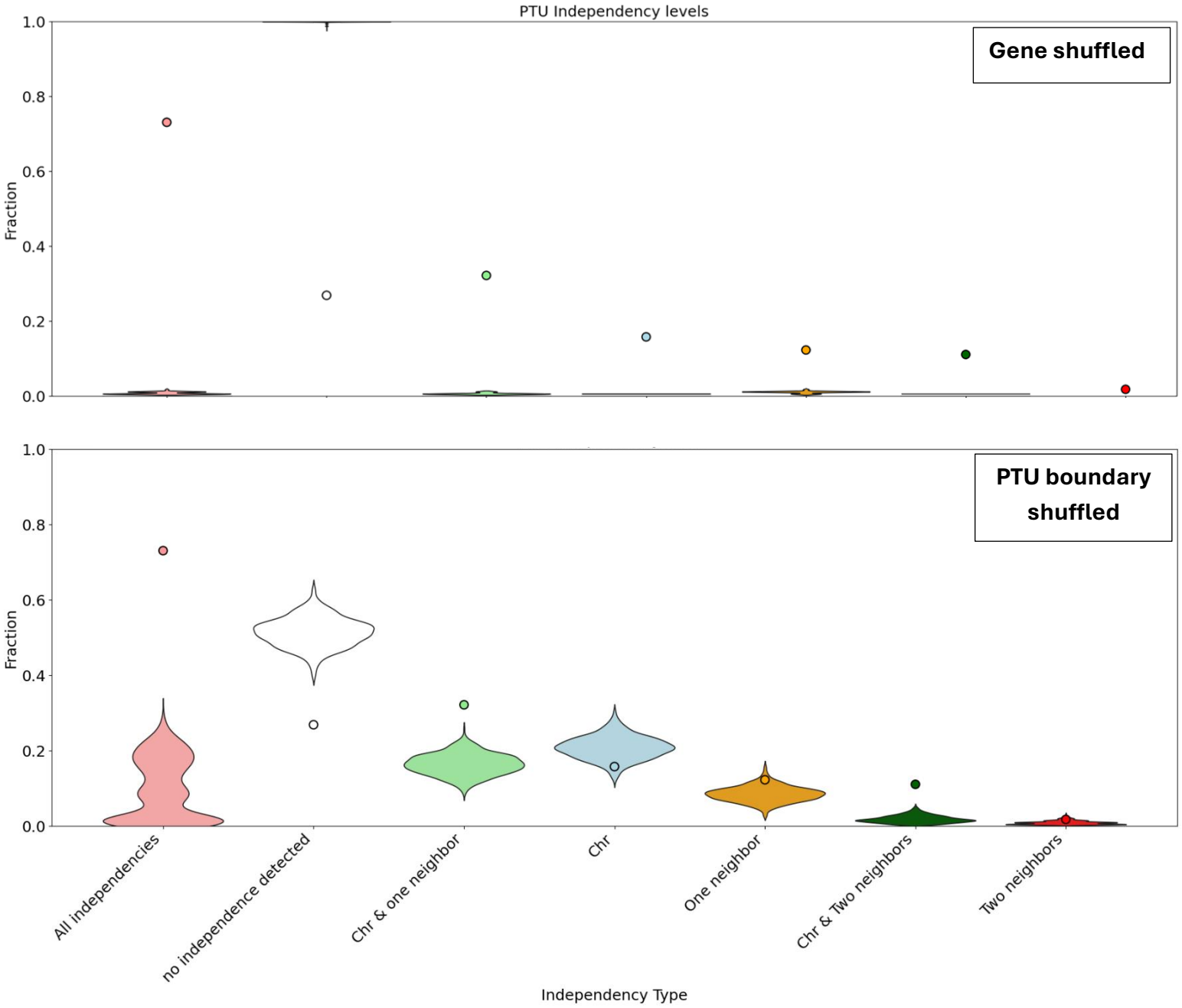


Figure 4C: **Significance of detected PTU independence relative to randomization.** PTU independence was evaluated against two randomization controls. In the first, genes were shuffled within each chromosome, reassigning each gene to a new genomic position and therefore potentially to a different PTU on the same chromosome (**top**). In the second, PTU boundaries were shuffled while preserving the empirical distribution of PTU lengths (measured by gene count) (**bottom**). Each independence category is shown separately using the same color scheme as in Figure 4B; the two leftmost groups summarize, respectively, the total fraction of PTUs classified as independent (any category) and the fraction with no independence detected. Violin plots show the distributions obtained from randomized datasets, and the circle denotes the observed value from the real data.

Temperature responsiveness correlates with GC content of the beginning of the PTU

Classical Motif Enrichment

Following the identification of PTUs that exhibit independent regulatory behavior across conditions, we investigated whether co-regulated PTUs share common sequence features that could account for coordinated transcriptional activation (**Figure 5A**). Specifically, we tested whether independently upregulated or downregulated PTUs contain enriched sequence motifs consistent with promoter-like elements or alternative regulatory sequences.

Across all tested parameter combinations, detected motifs showed limited statistical support. Most enriched motifs exhibited modest to borderline significance scores, ranging from 3.3×10^{-3} to 4.9×10^{-2}. The only hit with highly significant motif (E-value $\approx 7 \times 10^{-21}$) corresponded to the telomeric repeat sequence (**Figure 5B**), reflecting known sequence composition rather than regulatory specificity. No robust, reproducible sequence motif consistent with canonical eukaryotic promoters was identified.

Given the lack of strong motif enrichment and the established absence of canonical promoter architecture in trypanosomatids, these results suggest that PTU-specific regulation is unlikely to be mediated by discrete promoter motifs in the conventional sense.

GC Content as a Candidate Regulatory Feature

We therefore considered the possibility that transcriptional regulation operates through broader sequence properties rather than discrete motifs. Analysis of GC content revealed that PTUs upregulated under high-temperature conditions exhibit slightly higher GC content compared to both non-responsive PTUs and PTUs downregulated under the same condition (**Figure 5C**). This analysis focused on sequences spanning the segment from TSS till the fourth downstream genes and was performed using a sliding-window approach to capture local compositional variation.

Notably, the GC enrichment was not confined to upstream regions but was most pronounced within the early portion of the PTU itself. This observation, together with the GC pattern identified in the meta-PTU analysis showed the first PTU genes' pronounced deviation in GC content relative to downstream genes. This observation suggests that the relevant regulatory signal is embedded within the transcribed region rather than in upstream sequences where canonical promoter elements would typically be expected.

To further examine this possibility, we analyzed the relationship between gene GC content and transcriptional behavior, focusing on both absolute expression levels and transcriptional responses to elevated temperature. Because deviations in GC content are most prominent at the beginning of PTUs, we tested the correlation between GC content and temperature-induced expression changes as a function of gene position within the PTU.

As shown in **Figure 5D (top)**, cumulative correlation analysis revealed a strong positive correlation between GC content and temperature response when progressively including genes from the beginning of the PTU, with the highest correlation observed when only the earliest genes were considered. In contrast, sliding-window analysis demonstrated that this correlation is strongest for windows encompassing the first few genes and rapidly diminishes for downstream positions, approaching zero or becoming negative further along the PTU. These results indicate that the association between GC content and temperature responsiveness is largely restricted to the initial genes of PTUs. Based on these observations, subsequent analyses focused on the first four genes of each PTU, which consistently exhibited the strongest correlation between GC content and transcriptional response to temperature.

If GC content contributes to transcriptional regulation, it would be expected to correlate negatively with basal expression levels, due to thermodynamic constraints on DNA strand separation and elongation through GC-rich regions. Conversely, it would be expected to correlate positively with transcriptional responses to elevated temperature stress, because higher temperature reduces the energetic barrier for melting and polymerase progression, thereby preferentially relieving repression of GC-rich genes. Consistent with this expectation, we observed a significant negative correlation between GC content and basal expression levels under standard growth conditions (P30, 27 °C), indicating that genes with higher GC content tend to exhibit lower steady-state expression (**Figure 5D, middle**). In contrast, GC content showed a positive correlation with transcriptional response to elevated temperature, as measured by $\log_2$ fold change between 27 °C and 35 °C (Pearson r = 0.257, p = 1.93×10^{-14}; **Figure 5D, bottom**). These opposing correlations suggest that GC-rich genes are transcriptionally constrained under basal conditions but preferentially induced upon temperature increase.

To further explore the mechanistic basis of this relationship, we examined the association between GC content and RNA polymerase II occupancy using ChIP-seq data from *Trypanosoma brucei*. High

GC content is predicted to increase the energetic cost of transcription elongation, potentially leading to reduced polymerase progression and lower transcriptional output under basal conditions. In this framework, elevated temperature may partially alleviate this constraint by providing DNA melting conditions needed to open the double helix for transcription.

Consistent with this model, sliding-window analysis along chromosomes revealed a strong positive correlation between GC content and RNA polymerase II occupancy (Spearman $\rho = 0.606$, $p < 1 \times 10^{-16}$; **Figure 5E**). Together, these results support a model in which sequence composition modulates transcriptional dynamics and contributes to PTU-specific regulatory responses, particularly under temperature stress.

Figure 5

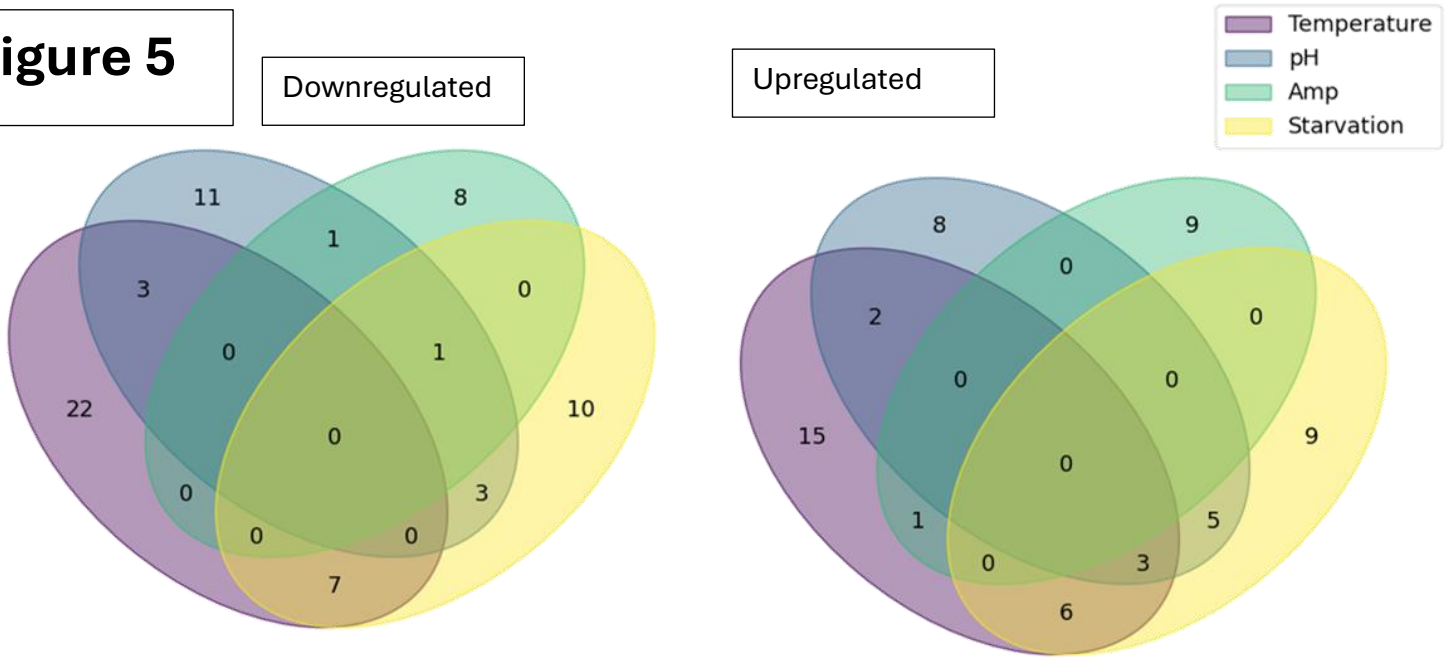


Figure 5A: Shared independent PTUs across conditional responses. Venn diagrams summarize overlap and condition-specific sets of independent PTUs. The left diagram shows PTUs classified as downregulated relative to their genomic context (significant in at least one comparison to neighboring PTUs and/or the chromosome background), and the right diagram shows the corresponding set of upregulated PTUs. Colors denote the condition.

Figure 5B: Motif captures the telomeric repeat sequence. Sequence logo representation of the enriched motif, which corresponds to the canonical telomeric repeat TTAGGG. Letter height reflects the information content at each position (in bits) and is proportional to nucleotide frequency in the aligned instances; larger letters indicate higher confidence/stronger conservation of that base at the given position. Sequences were extracted relative to inferred TSSs. Multiple parameters were systematically explored: The genomic location of candidate regions (upstream of the TSS or spanning the TSS and the first four downstream genes), as well as the length of the extracted sequences (100 bp, 500 bp, 1 kb, 2.5 kb, 5 kb, 7.5 kb, and 10 kb). To evaluate condition specificity, different target and background sets were defined, including: (i) all PTUs upregulated under a given condition, (ii) PTUs upregulated exclusively under a specific condition, and (iii) PTUs repeatedly upregulated across multiple comparisons of the same condition type. Background sets consisted of PTUs that were either not up/ down regulated or non-responsive. MEME was run in discriminative mode, using background sequences as negative controls to enrich for motifs specific to the target PTU sets. Analyses were performed on DNA sequences (-dna) using the differential enrichment objective function (-objfun de) with a zero-order Markov background model (-markov_order 0). Motifs were searched on both strands (-revcomp) under the assumption of zero or one motif occurrence per sequence (-mod anr). For each run, up to five motifs were identified (-nmotifs 5), with motif widths constrained between 5 and 50 bp (-minw 5, -maxw 50). Statistical significance was assessed using an E-value threshold of 0.05 (-evt 0.05).

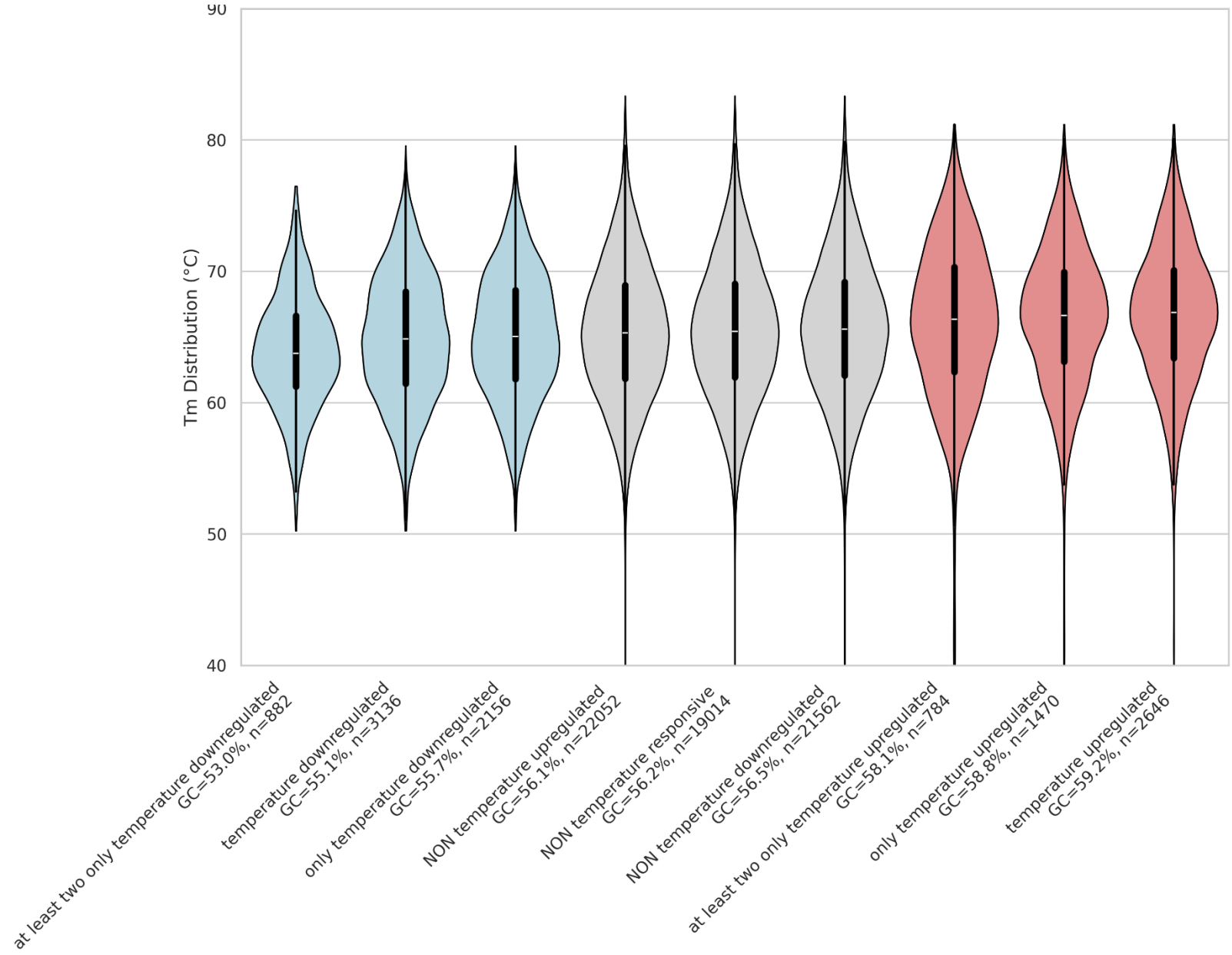
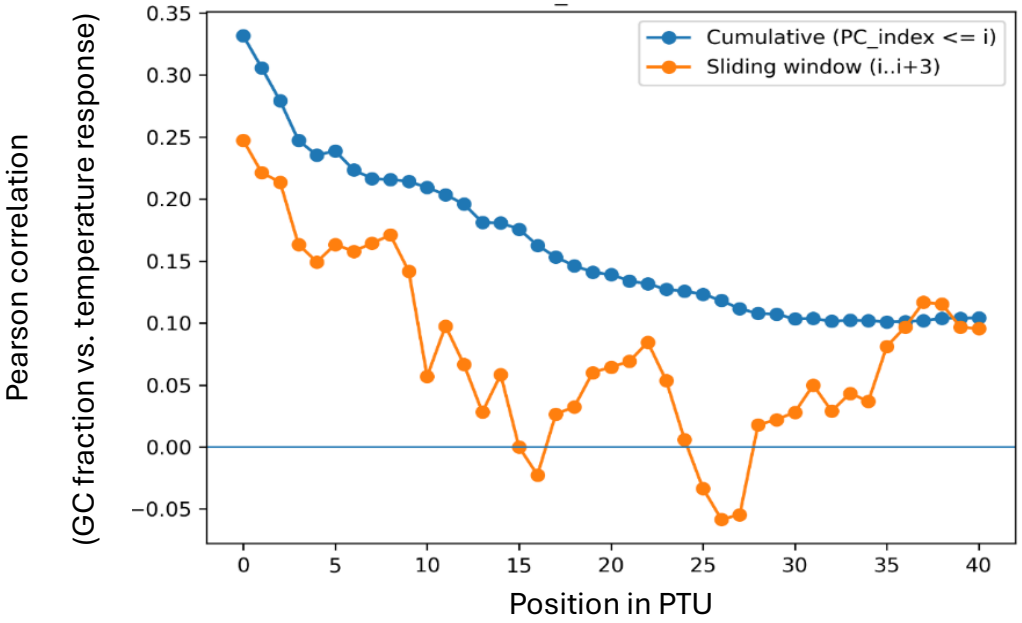
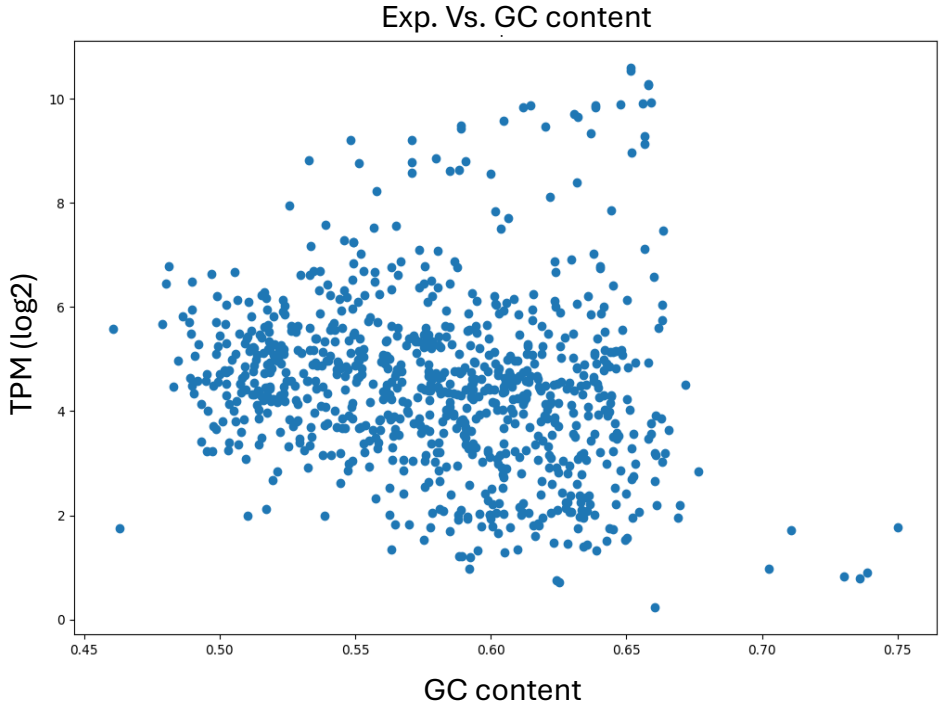


Figure 5C: Melting temperature profiles at PTU starts stratified by temperature-response specificity. Violin plots show the distribution of DNA melting temperature (Tm) values computed using a sliding-window approach across PTU start regions. For each PTU, the analyzed sequence spans up to 1,000 bp upstream of the first nucleotide of the fourth gene. PTUs were grouped into three temperature-response specificity classes: “at least two only temperature”, PTUs showing the indicated response in at least two comparisons and exclusively under temperature conditions (i.e., not detected with the same response in any other condition); “only temperature”, PTUs showing the indicated response only under temperature (even if detected in a single comparison); and “just up/downregulated”, PTUs with the indicated response that may also show the same response under other conditions. Colors denote response direction and background: light blue, downregulated; red, upregulated; gray, background PTUs not classified as temperature-responsive in the corresponding category. Melting temperatures (Tm) were computed from FASTA-formatted sequence sets using the nearest-neighbor model implemented in Biopython (Bio.SeqUtils.MeltingTemp.Tm_NN). Tm was estimated using a sliding-window approach to obtain a robust within-sequence distribution rather than a single value per sequence. For each FASTA file, each sequence was scanned using windows of fixed length (15 bp) advanced by a fixed step size (5 bp). For every window, Tm was calculated and aggregated into a per-sequence distribution of window-level Tm values. For each sequence, the mean and median of its window-level Tm distribution were computed; file-level summary statistics were then derived as the mean of per-sequence means (overall mean Tm) and the median of per-sequence medians (overall median Tm). In parallel, GC content was calculated per sequence as the fraction of G/C bases among all nucleotides, and file-level GC content was summarized as the mean across sequences.





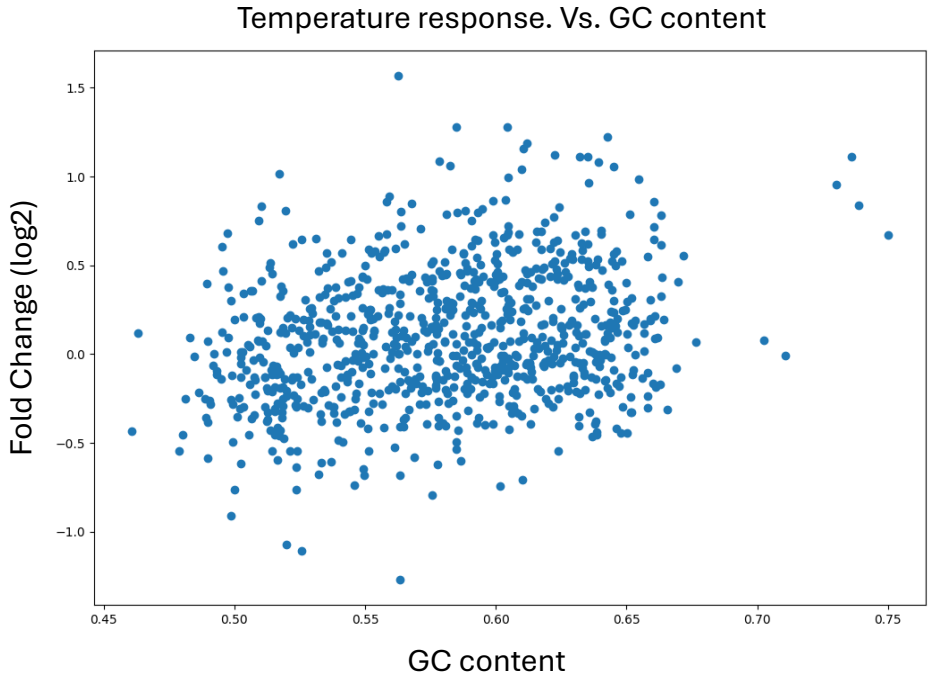


Figure 5D: Top: GC-temperature response correlation as a function of position along the PTU. Curves show the Pearson correlation between gene GC content and the transcriptional response to elevated temperature (log2 fold change). The cumulative curve (blue) reports the correlation calculated using genes 1 through i (e.g., at i = 10, the correlation is computed across the first 10 genes of each PTU). The sliding-window curve (orange) reports correlations computed in a moving window of four consecutive genes (genes i to i+3). Both analyses indicate that the strongest correlation is concentrated at the beginning of PTUs, with the peak signal arising from the first four genes.

Middle: Raw expression as a function of GC content. Scatter plot of baseline gene expression (log2 TPM) at 27°C (passage 30) versus gene GC content. Each point represents a gene; only genes occupying positions 1–4 within a PTU were included.

Bottom: Temperature response as a function of GC content. Same analysis as the middle panel, but the y-axis shows the temperature response as log2 fold change (35°C versus 27°C) at passage 30.

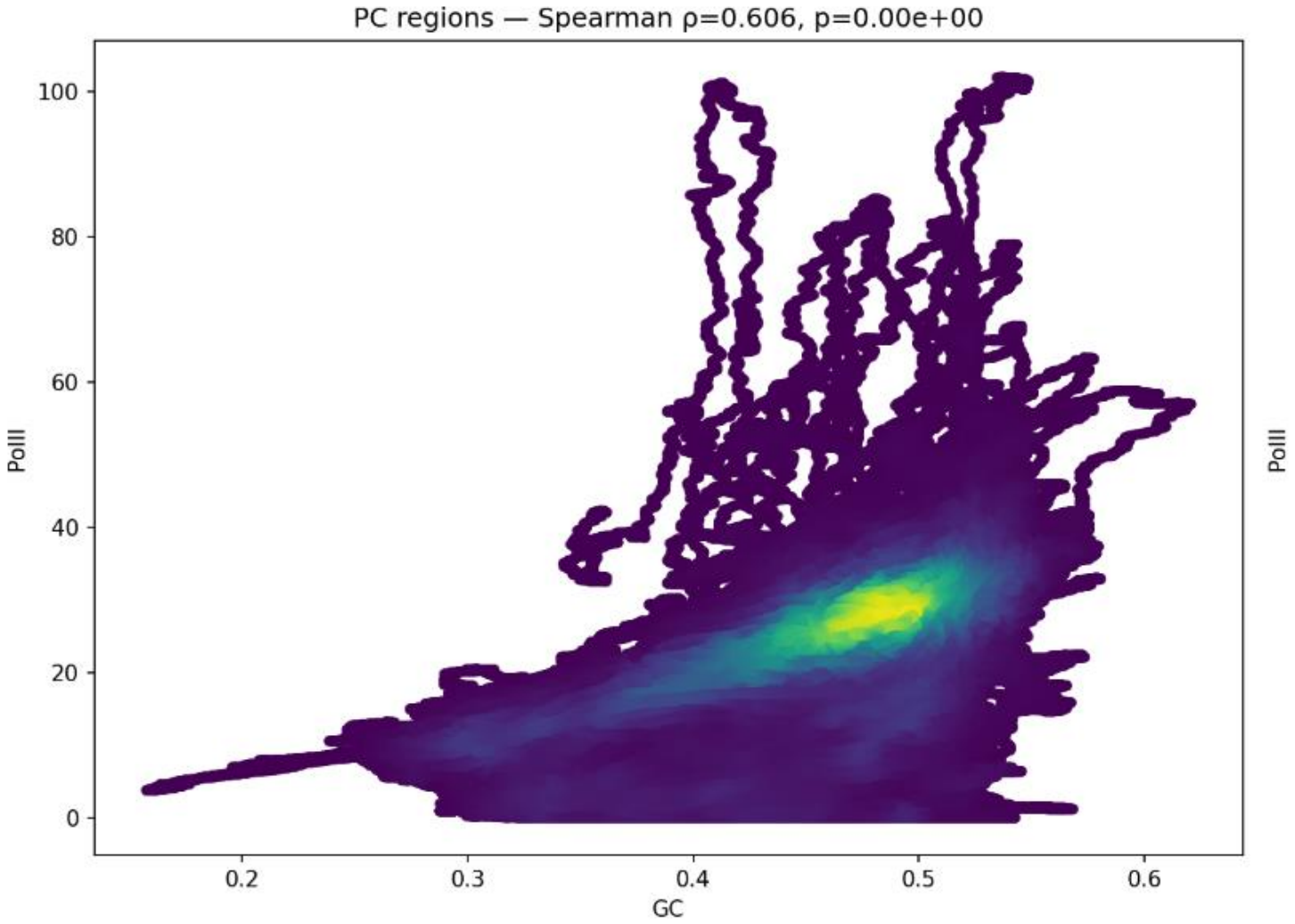


Figure 5E: GC content is positively associated with Pol II occupancy within polycistronic regions in T. brucei. Each point represents a centered sliding-window measurement (GC fraction versus log2 Pol II) and is colored by local point density. Spearman ρ=0.606, p-value= $p < 1 \times 10^{-16}$).

Part 2: *Leishmania's* expression regulation by translation efficiency

Codon usage is not good predictor for gene functional classification

Codon usage refers to the frequency with which synonymous codons are used within protein-coding genes. Codon usage can be quantified either as raw codon frequencies or after normalization by amino acid, such that the frequencies of all codons encoding the same amino acid sum to one. Both representations were used in this analysis.

Codon usage profiles were calculated for all genes in *L. donovani* and subjected to principal component analysis (PCA) to assess whether genes that belong to shared expression or functional groups cluster according to codon composition. Using raw codon frequencies, the first and second principal components explained approximately 80% and 5% of the total variance, respectively; however, this variance was driven by a small number of outliers, while the vast majority of genes were densely clustered within a much narrower range. In contrast, amino acid–normalized codon usage explained approximately 18% and 7% of the total variance for PCA1 and PCA2, respectively, with most data points distributed across this range, indicating that codon usage contributes to measurable but relatively limited variability across genes.

To examine whether codon usage is associated with condition specificity, PCA projections were colored according to expression levels under different experimental conditions. Expression was evaluated using transcript abundance (TPM) or rank-normalized values between 0 and 1. To assess responsiveness, differential expression values ($\log_2$ fold change) were similarly used, either directly or after rank normalization between 0 and 1. This analysis was performed using both raw and amino acid–normalized codon usage matrices. Across all conditions and representations, no clear separation or clustering of genes based on expression level or fold change was observed, suggesting that codon usage alone does not strongly predict condition-specific transcriptional responses.

Codon usage PCA was additionally performed across all species included in the evolutionary context analysis, with genes colored by functional pathway annotation. In most cases, pathway-based coloring did not reveal distinct codon usage–based clustering. However, two notable patterns were observed (**Figure 6A**). First, genes encoding ribosomal proteins consistently clustered at the periphery of the PCA space, suggesting that they occupy a highly optimized codon usage 'niche'.

Second, in a subset of species, genes associated with disease-related functions formed a distinct cluster, a pattern that may be consistent with acquisition through horizontal gene transfer. Notably, the disease-related annotation assigned to these genes was “Trypanosomiasis.” This annotation should be interpreted with caution, as it is based on sequence similarity to genes from other species with established functional annotations. In practice, “Trypanosomiasis” primarily refers to Trypanosoma-associated disease; however, given the shared gene repertoire and functional overlap across trypanosomatids, it is reasonable to interpret this signal more broadly as enrichment for disease-associated functions rather than as a Trypanosoma-specific designation.

tRNA Pool Composition

To evaluate the potential contribution of tRNA availability to translation efficiency, tRNA gene content was predicted using tRNAscan across the analyzed species. Several species exhibited a surprisingly small number of tRNA genes, consistent with previous reports[18]. For example, only 82 and 70 tRNA genes were detected in *Leishmania major* and *Trypanosoma brucei*, respectively. These values are more reminiscent of bacterial tRNA gene counts than those typically observed in eukaryotes.

For some species, genome assembly quality was relatively low, which may affect the accuracy of tRNA gene detection. Consequently, tRNA gene counts for these species should be interpreted with caution.

Analysis of anticodon distributions revealed broad similarity across species, with modest variation in the abundance of specific anticodons (**Figure 6B**). In addition, several species appeared to contain “disallowed” anticodons, defined as anticodons not expected based on canonical decoding rules[32]. Further validation will be required to distinguish biologically meaningful signals from artifacts arising from incomplete or low-quality genome assemblies

Projecting tRNA copy number against the usage frequencies of their cognate codons revealed a modest correlation between supply (tRNA availability) and demand (codon usage). Across most species, the correlation coefficient was approximately 0.33, which is lower than values reported for some yeasts.[33]

As mentioned above, one of the known regulation mechanisms is *Leishmania* is altering the chromosome copy number. In addition to the change in coding genes dosage, it may also affect the tRNA genes of chromosome changed.

Most amino acids have similar trend of tRNA distribution across conditions, but for Alanine, there is a signal of possible conditional codon preference (**Figure 6D**). The anti-codon TGC appears relatively low in all conditions, but the two other anti-codons, AGC and CGC have small differences in untreated controls (mean of 5.80 and 5.47, respectively). For the low pH samples there is a slightly more from the CGC anti-codon (mean of 6.50 vs. 6.05 of CGC), while in the elevated temperature the AGC anti-codon has higher count compared to CGC (mean of 7.01 and 4.39, respectively).

Figure 6

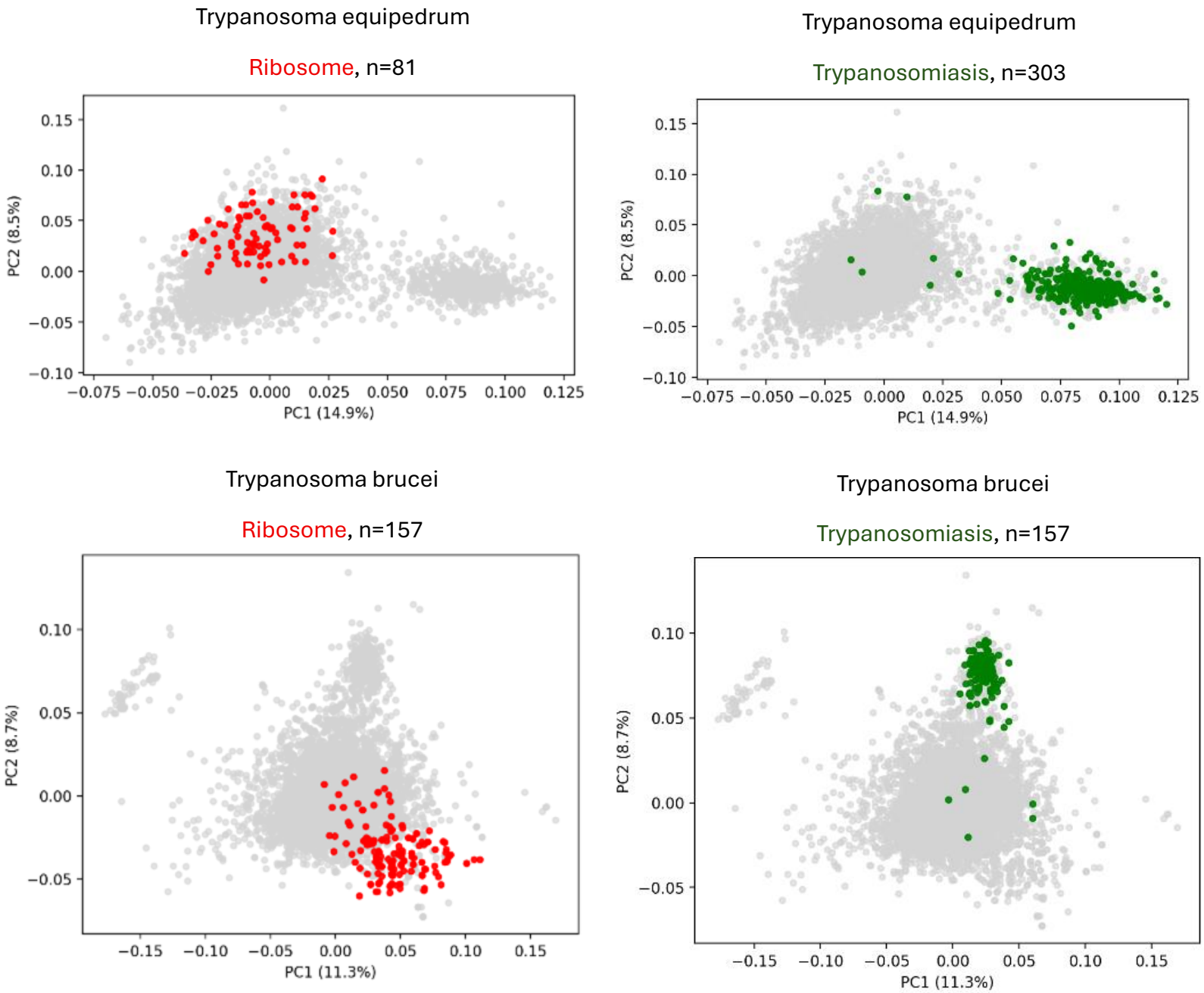
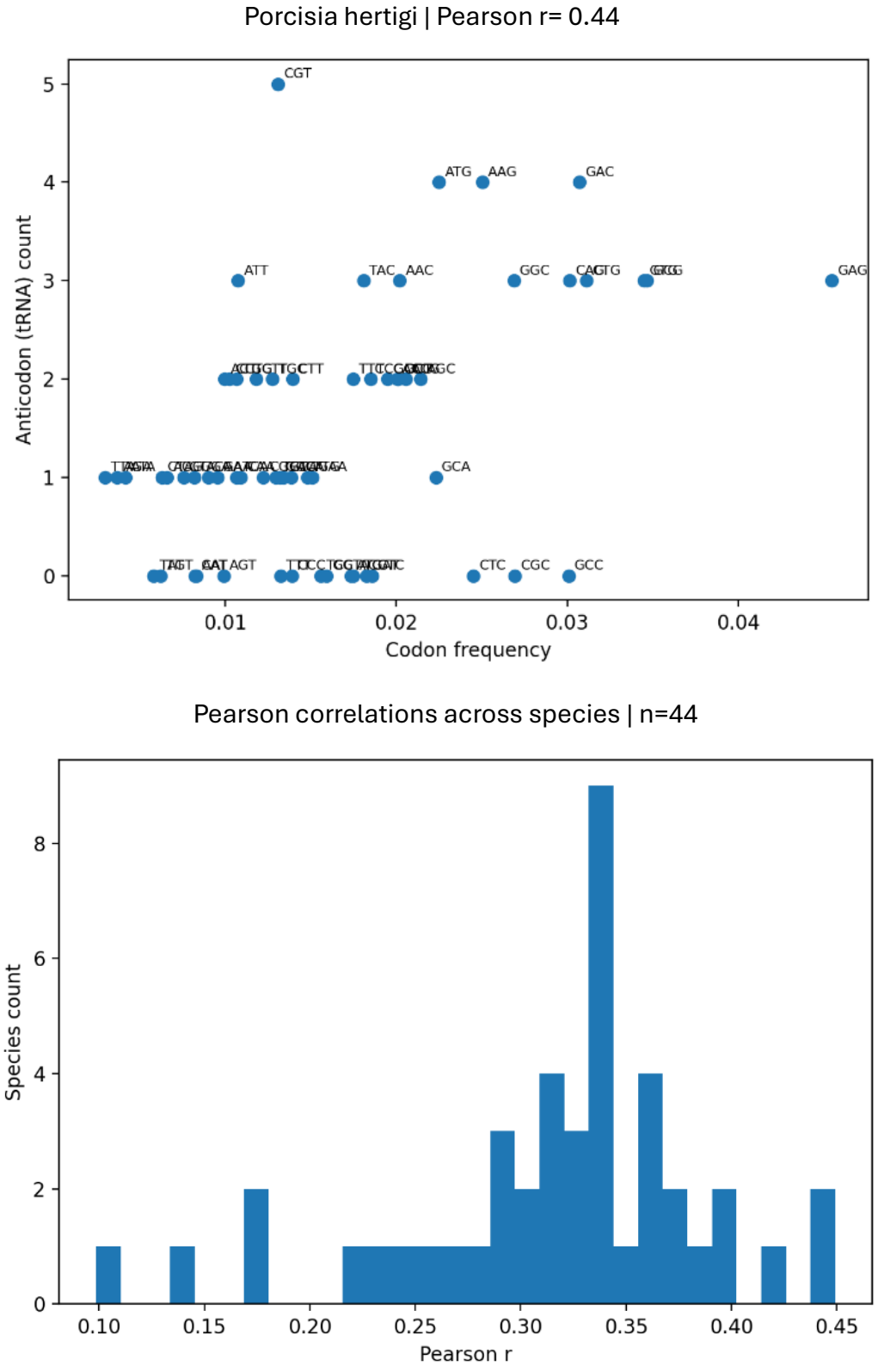


Figure 6A: Ribosomal and trypanosomiasis-associated genes exhibit distinct codon-usage profiles. The four panels show the distribution of genes based on codon-usage frequencies. The top two panels correspond to *Trypanosoma equiperdum*, and the bottom two to *Trypanosoma brucei*. Ribosome-related genes are highlighted in red, and trypanosomiasis-associated genes are highlighted in green. Genes with CDS length <150 nt were excluded. Gene – pathway relations were determined by previous analysis in our group by sequence similarity.

Figure 6B: tRNA gene copy-number landscape across trypanosomatids. Heatmap shows tRNA gene copy numbers per anticodon across all analyzed species (rows) and sense-anticodons (columns). Cell color encodes copy number. Anticodon columns are hierarchically clustered using Euclidean distance and Ward linkage. Row order is defined by hierarchical clustering computed on the subset of species overlapping the tAI dataset. tRNAscan-SE output files were p filtered by a minimum prediction score threshold (≥50).. As a basic validity control, species were excluded if the

Figure 6C: Correlation between codon usage and matching tRNA anticodon abundance across species. For each species, the association between codon frequencies and the copy numbers of the corresponding tRNA anticodons was quantified using Pearson correlation (example shown in the top panel). Pearson *r* values from all species were then compiled and visualized as a histogram in the bottom panel, summarizing the distribution of codon–tRNA coupling strengths across the dataset.



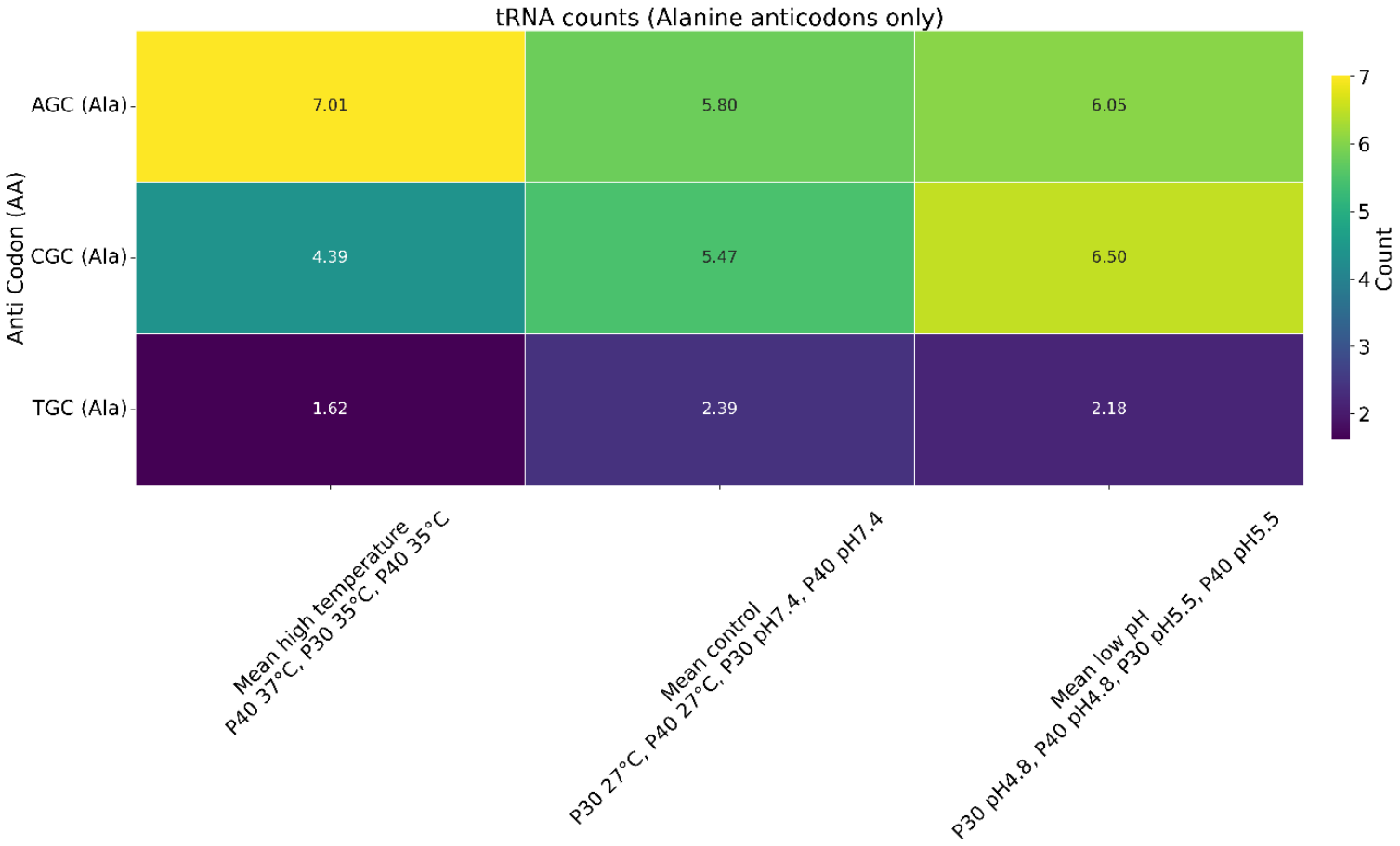


Figure 6D: Alanine tRNA abundance across conditions. Heatmap shows mean tRNA counts for alanine (Ala) anticodons (rows) after collapsing samples into three condition groups (columns): high temperature (P40 37°C, P30 35°C, P40 35°C), control (P30 27°C, P40 27°C, P30 pH 7.4, P40 pH 7.4), and low pH (P30 pH 4.8, P40 pH 4.8, P30 pH 5.5, P40 pH 5.5). Values represent the mean across the listed replicates within each group. tRNA loci were identified by integrating genome annotation (GFF) and tRNAscan-SE predictions. tRNAscan-SE output was filtered and only high-confidence predictions were retained (score ≥ 50).

Genome Assembly Validation and tAI Reliability

Following the examination of codon usage and tRNA pool composition as individual contributors to translation efficiency, we next employed the tRNA adaptation index (tAI)[26] to integrate both factors into a single quantitative measure of predicted translation efficiency. The tAI scores genes based on their codon composition and the availability of corresponding tRNAs.

Because tRNA gene content was inferred using tRNAscan, which depends on the quality of the underlying genome assembly, inaccurate or incomplete assemblies may lead to erroneous tRNA detection and consequently unreliable tAI estimates. To assess the reliability of genome assemblies and the validity of inferred tAI values, we examined the relationship between tAI scores and codon entropy for each species, as done before too[34].

Codon entropy is expected to be negatively correlated with tAI, such that genes with higher predicted translation efficiency (higher tAI) exhibit lower codon entropy. Consistent with this expectation, several species showed a clear negative correlation between tAI and codon entropy. However, the strength and direction of this relationship varied across species, with some deviating from the expected trend (**Figure 7A**).

Species-Specific Patterns of tAI in Protein Families

To examine whether translation efficiency exhibits species-specific patterns, we defined protein families across all analyzed species by performing an all-versus-all BLAST of coding sequences. Protein families were constructed based on sequence similarity, and for each species, the tAI of each protein family was calculated as the mean tAI of its member genes.

Hierarchical clustering of species based on protein family–averaged tAI profiles revealed a pattern that closely mirrors the phylogenetic relationships among the species (**Figure 7B**). *Leishmania* species clustered together, as did *Trypanosoma* species, while an intermediate group comprising *Porcisia hertigi*, *Endotrypanum monterogeii*, *Crithidia fasciculata*, *Blepharomonas ayalai*, and *Angomonas deanei* formed a distinct subgroup. *Bodo saltans* clustered separately, consistent with its role as an outgroup. These results indicate that translation efficiency, as captured by tAI, reflects evolutionary divergence among trypanosomatids.

Clustering protein families based on their vectors of average tAI values across species further revealed both shared and lineage-specific translation efficiency patterns (**Figure 7C**). One cluster, conserved across all species and characterized by high tAI values, contain ribosomal proteins, consistent with strong translational optimization. Additional clusters displayed lineage-specific patterns, including clusters with elevated tAI values specific to all *Trypanosoma* species or to subsets thereof, as well as a cluster specific to *Bodo*. Overall, this clustering highlights differential translational adaptation of protein families across species. Broader functional annotation of these protein families would enable further biological interpretation, such as associations with parasitic versus free-living lifestyles or adaptation to distinct ecological niches.

Condition-Dependent Changes in tAI during Experimental Evolution

To assess whether translation efficiency changes during experimental adaptation, tAI values were calculated and compared across conditions in the laboratory evolution experiment. While codon usage is constant across conditions, the tRNA pool can vary as a consequence of chromosome copy number changes. Overall, tAI values were highly correlated across conditions, indicating that global translation efficiency remained largely stable during experimental evolution.

However, under temperature-adapted conditions, four genes deviated from the otherwise tight correlation along the y = x line (**Figure 7D**). Specifically, the genes exhibited markedly increased tAI values relative to the untreated control as a result of chromosome somy and hence tRNA gene copy number changes. These genes were annotated as trans-sialidases, enzymes specific to trypanosomatids that transfer sialic acid from host glycoconjugates to the parasite surface, thereby contributing to immune evasion.

Notably, trans-sialidase genes are known to be more highly expressed in the amastigote form (**Figure 7E**), suggesting that increased translation efficiency of these genes under elevated temperature conditions may reflect adaptive regulation associated with host-related environments. This analsysis suggests that not only that these genes are induced at the RNA level in the amastigote, but that a tRNA that is specific to their unique codon usage gains in copy in this condition.

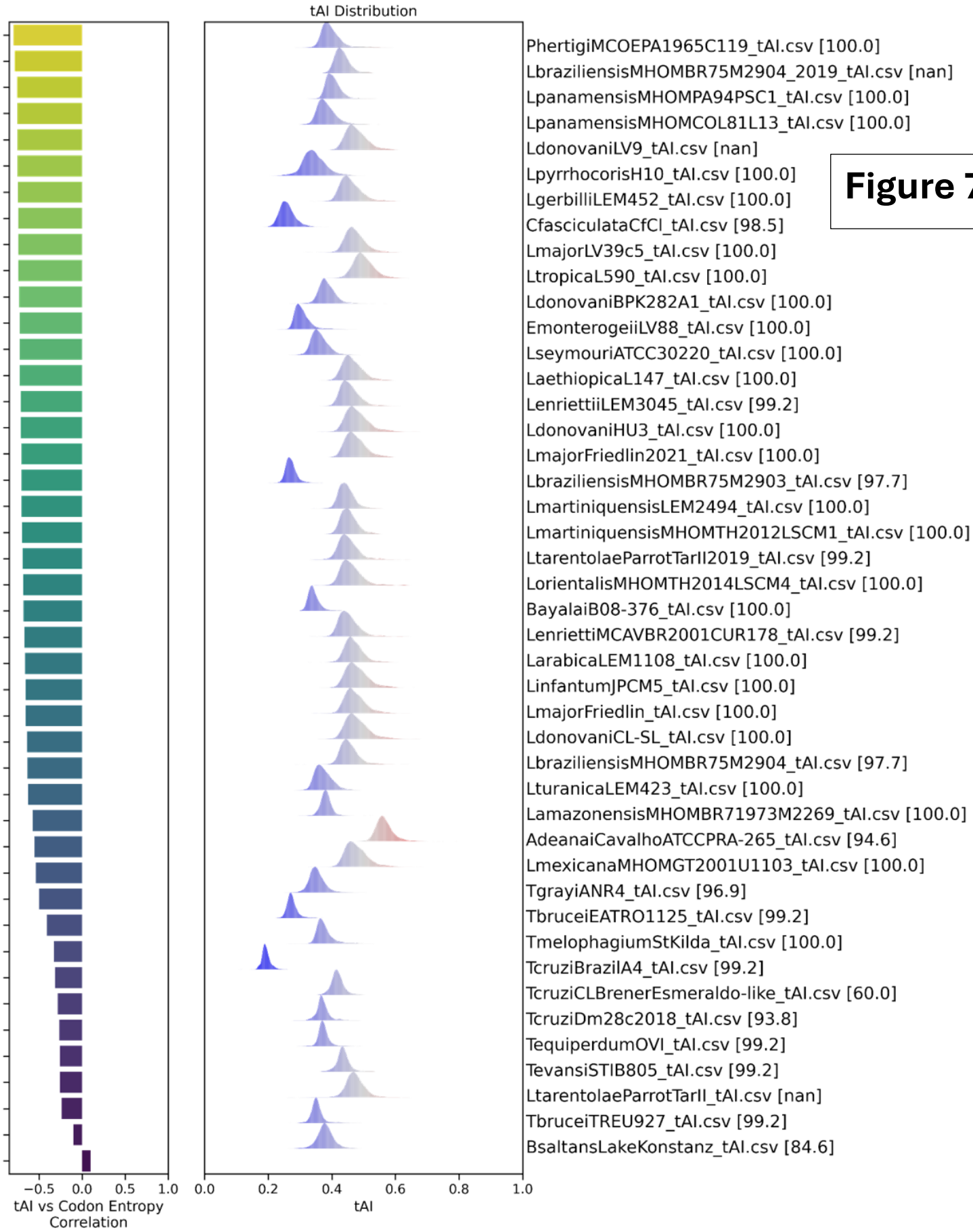


Figure 7

Figure 7A: Species-level coupling between translational adaptation and codon-usage complexity. For each species, tRNA adaptation index (tAI) values were and correlated with codon-usage entropy across coding sequences (Pearson correlation). The left panel summarizes the resulting species-specific correlation coefficients (tAI versus codon entropy). The right panel shows the distribution of gene-level tAI values for each species. Bracketed values next to species names indicate BUSCO genome completeness scores used for filtering/quality control.

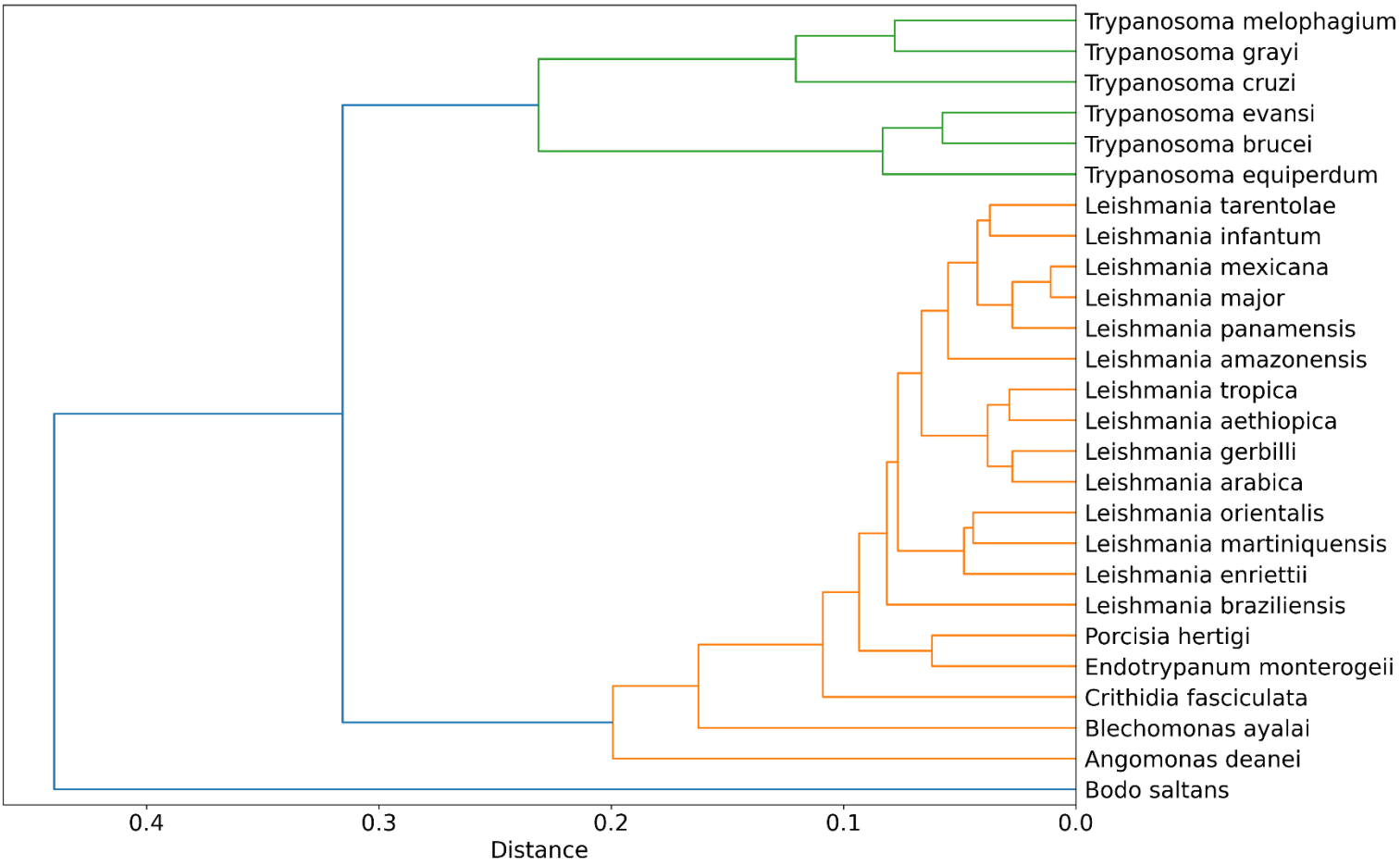


Figure 7B: tAI-profile clustering recapitulates the species relationships shown in Figure 1C. Data values indicate the mean tRNA adaptation index (tAI) of genes assigned to each protein family (orthogroup; defined by sequence similarity) in each species. Species were hierarchically clustered using a correlation distance metric and complete-linkage clustering.

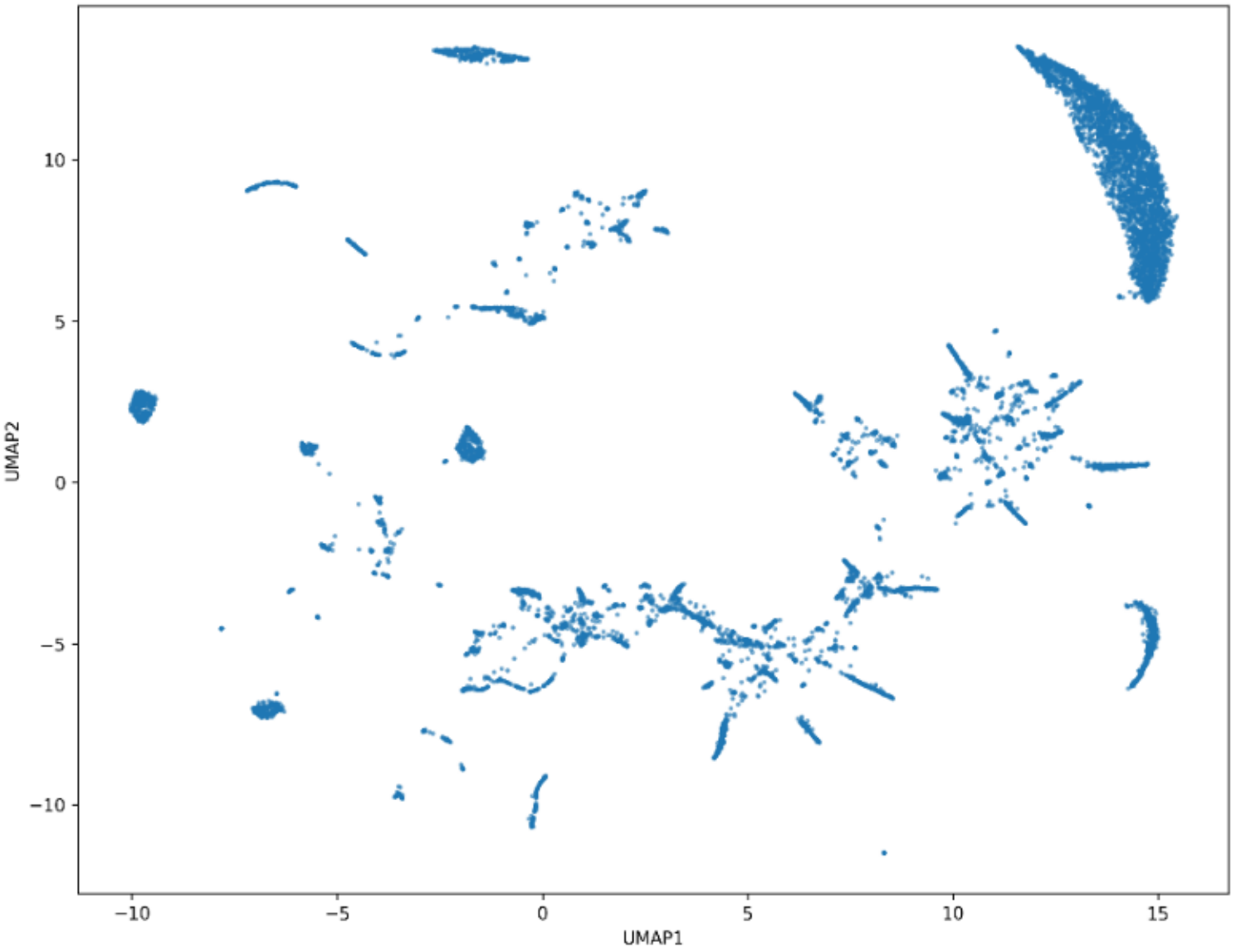


Figure 7C: UMAP projection of protein families based on genome-wide tAI patterns. Each point represents a protein family (orthogroup) defined by sequence similarity. For each family, the feature vector comprises the standardized mean tRNA adaptation index (tAI) across species (mean tAI calculated over all genes assigned to that family in each species; species tAI scores z-scored to make species comparable). UMAP was applied to this family-by-species matrix (n_neighbors = 30, min_dist = 0.1, Euclidean metric), yielding a two-dimensional embedding (UMAP1/UMAP2) that groups protein families with similar cross-species tAI profiles.

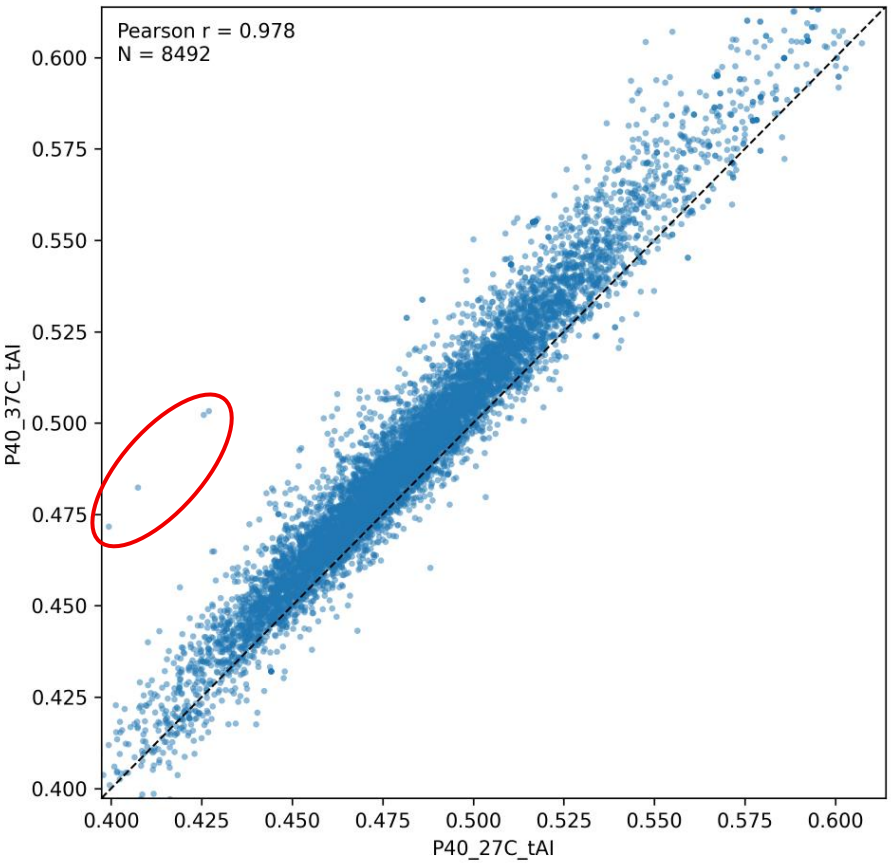
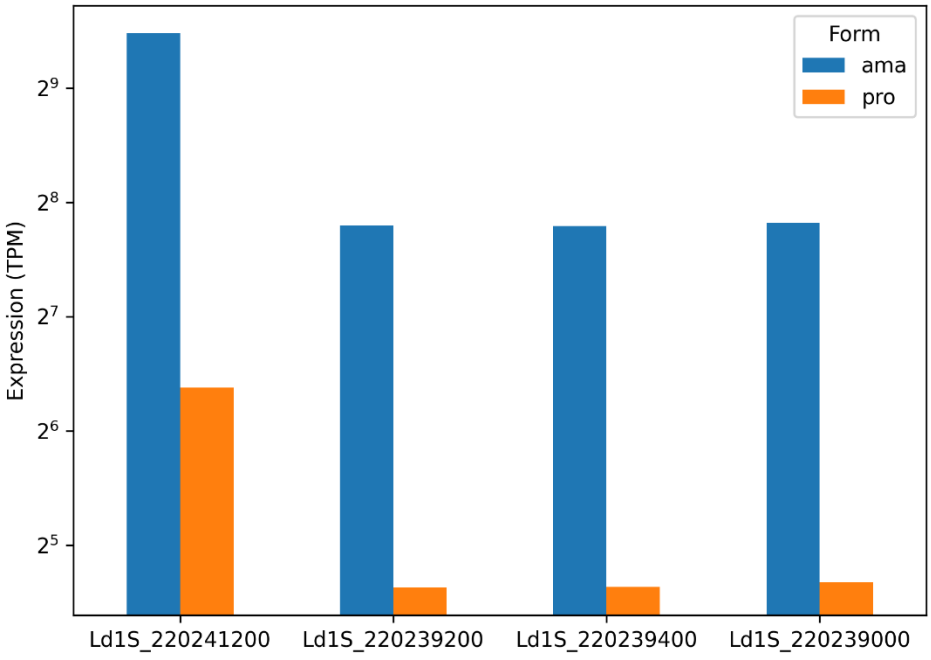


Figure 7D: Gene-wise tAI comparison between P40 27°C and P40 37°C. Scatter plot compares per-gene tRNA adaptation index (tAI) values at passage 40 measured under control temperature (27°C; x-axis) versus elevated temperature (37°C; y-axis). Each point represents a gene. points above the line have higher tAI at 37°C than at 27°C, and points below have lower tAI at 37°C. Pearson correlation coefficient (*r*) and the number of genes (N) are reported in the panel.

Red circle mark the four extreme outlier genes that deviate from the general y = x trend, exhibiting higher tAI at elevated temperature (37°C) relative to 27°C. These four genes are annotated as trans-sialidases.

Figure 7E: Expression of trans-sialidase genes across life-cycle forms. Bar plot shows the expression levels (TPM; log2 scale) of four trans-sialidase genes (Ld1S_220239400, Ld1S_220241200, Ld1S_220239000, Ld1S_220239200) across two life-cycle forms. For each gene, “ama” (for amastigote) and “pro” (for promastigote) represent the mean TPM across the corresponding replicates (ama.1/ama.2 and pro.1/pro.2). Legend indicates the form.



Discussion

In this study, we investigated whether polycistronic transcription units (PTUs) in *Leishmania* function solely as structural features of genome organization or also serve as independent regulatory entities. Our analysis revealed several consistent and interrelated findings. First, multiple molecular features-including GC content, mRNA stability, translation efficiency, transcript abundance, and RNA polymerase II occupancy-exhibit conserved positional patterns along PTUs, and functionally related genes tend to occupy similar positions within PTUs. Although these correlations are moderate, they support a model in which PTU architecture contributes to functional coherence, enabling groups of genes to be co-regulated in response to environmental cues. In the absence of canonical gene-specific transcriptional regulation, these positional patterns suggest that evolutionary selection may have acted on gene positioning within PTUs. Specifically, genes requiring higher expression may be preferentially located at earlier PTU positions, thereby exploiting the elevated transcriptional output characteristic of PTU beginnings. This dosage–position relationship is further supported by the coordinated positioning of enzymes within the same metabolic pathway and of subunits of heteromeric protein complexes. Such coordinated positioning could promote comparable expression levels among functionally linked genes, thereby supporting efficient metabolic flux and appropriate stoichiometry during complex assembly. Second, a substantial fraction of PTUs responds independently to conditional and evolutionary perturbations, displaying transcriptional behaviors that differ from neighboring PTUs and from chromosome-wide trends. Third, PTU independence is observed across different experimental platforms, including RNA-seq and PRO-seq, and is partially conserved across trypanosomatid species. Finally, randomization controls demonstrate that these patterns depend on authentic PTU boundaries rather than arbitrary genomic segmentation or random reassignment of genes within chromosome. Together, these results support the existence of PTUs as biologically meaningful units of gene regulation.

PTUs as an Intermediate Regulatory Layer

Gene regulation in *Leishmania* is traditionally described as operating at two principal levels: chromosome-wide regulation through changes in copy number, and gene-specific regulation mediated predominantly by post-transcriptional mechanisms. The polycistronic organization of protein-coding genes and the apparent absence of canonical promoters have reinforced the view

that transcription initiation plays a limited role in regulating individual genes. However, the findings presented here indicate that this framework is incomplete.

Our data supports a model in which PTUs constitute an intermediate regulatory layer between whole-chromosome dosage effects and gene-level post-transcriptional control. PTUs can exhibit coordinated transcriptional responses that are neither uniformly chromosome-wide nor entirely gene-specific. This intermediate scale of regulation allows *Leishmania* to fine-tune gene expression in response to environmental challenges while maintaining a polycistronic transcriptional architecture.

Evidence from PRO-seq analysis in *Leishmania major* indicates that PTU-level differences are detectable at the level of nascent transcription, prior to the influence of post-transcriptional processing. The coherence of transcriptional output within individual PTUs, coupled with consistent differences between adjacent PTUs, suggests that transcription initiation or early elongation rates differ among PTUs. These findings argue against models in which expression patterns arise exclusively from post-transcriptional regulation and instead point to regulated transcriptional activity at the level of individual PTUs.

In addition to differences in rates of nascent transcription, differential gene expression across conditions also can exhibit PTU-dependent responses. Using conventional RNA-seq data, we show that some PTUs display statistically significant expression differences relative to their adjacent genomic regions. Together, these two forms of PTU independence raise the possibility that PTUs function as regulatory units that can be selectively modulated to tune gene expression and facilitate adaptation to changing conditions.

Sequence Composition and GC-Dependent Transcriptional Modulation

A notable feature of PTU architecture is the presence of conserved GC content gradients, particularly pronounced at the beginning of PTUs. The association between GC content, basal expression levels, and temperature-induced transcriptional responses suggests that sequence composition may influence transcriptional dynamics. High GC content is known to increase the energetic cost of DNA strand separation and transcription elongation, potentially resulting in reduced transcriptional efficiency under basal conditions. Elevated temperature may partially alleviate this constraint, leading to preferential activation of GC-rich PTUs under heat stress.

This pattern may reflect a transcriptional “ramp” of GC-rich genes whose expression is constrained under basal conditions but becomes more efficient at elevated temperatures.

This model is further supported by the observed correlation between GC content and RNA polymerase II occupancy. Together, these findings suggest that PTU-specific sequence composition may contribute to differential sensitivity to environmental conditions, providing a mechanism for selective PTU activation without reliance on canonical promoter motifs.

More speculatively, GC content could act as an intrinsic “thermosensitive” regulatory feature embedded within PTUs: GC-rich regions may suppress basal transcription by imposing higher energetic barriers to DNA melting and polymerase progression, while elevated temperature could preferentially relax these constraints and thereby amplify transcription specifically from GC-rich loci. At the same time, heat stress may reduce mRNA stability by destabilizing RNA secondary structures and accelerating transcript turnover. Notably, genome-wide coupling between mRNA production and degradation has been reported, implicating RNA polymerase II as a potential coordinating factor[35]. If both processes occur simultaneously, the resulting expression program could be highly dynamic-favoring selective induction of heat-relevant genes or generating transient bursts driven by increased synthesis coupled with faster decay—thereby providing an additional, sequence-encoded layer of regulation that operates without canonical promoter motifs.

Conservation and Divergence Across Trypanosomatids

Comparative analyses indicate that PTU independence is not unique to *Leishmania* but is also detectable in *Trypanosoma brucei*, albeit to a lesser extent. The lower prevalence of independently regulated PTUs in *T. brucei* may reflect differences in genome organization, such as the smaller number of chromosomes and higher PTU density, or differences in the nature of environmental stresses encountered by the parasite. Alternatively, the reduced signal may result from limitations in available datasets. These observations suggest that while PTU-level regulation is a conserved feature of trypanosomatids, its extent and regulatory significance may vary between species.

The results presented here support a revised model of gene regulation in *Leishmania*, in which polycistronic transcription units function as independent, adaptable regulatory entities. This intermediate layer of regulation bridges the gap between chromosome-wide dosage effects and

gene-specific post-transcriptional control, providing a flexible mechanism for coordinated transcriptional responses.

Future studies should aim to experimentally validate PTU-specific transcription initiation mechanisms, investigate the role of chromatin remodeling factors, and integrate PTU-level regulation with translational control mechanisms. Such efforts will be essential for developing a comprehensive understanding of gene regulation in kinetoplastid parasites and may reveal new vulnerabilities exploitable for therapeutic intervention.

Literature

1. Pareyn, M. *et al.* Leishmaniasis. *Nature Reviews Disease Primers 2025 11:1* **11**, 81- (2025).
2. Cutaneous leishmaniasis: Epidemiology and control - UpToDate. https://www.uptodate.com/contents/cutaneous-leishmaniasis-epidemiology-and-control.
3. Cecílio, P., Cordeiro-da-Silva, A. & Oliveira, F. Sand flies: Basic information on the vectors of leishmaniasis and their interactions with Leishmania parasites. *Commun. Biol.* **5**, 305–305 (2022).
4. Herwaldt, B. L. Leishmaniasis. *The Lancet* **354**, 1191–1199 (1999).
5. Leishmaniasis. https://www.who.int/news-room/fact-sheets/detail/leishmaniasis.
6. Aryal, S., Chauhan, U. & Pokhrel, S. Leishmaniasis: Global Epidemiology, Transmission Dynamics, and Integrated Control Strategies. *Int. J. Trop. Dis. Health* **46**, 93–109 (2025).
7. Chakravarty, J., Seth, A. & Sundar, S. Visceral leishmaniasis: Recent updates. *Annals of Medical Science & Research* **4**, S53–S59 (2025).
8. Baneth, G. & Solano-Gallego, L. Leishmaniasis. *Veterinary Clinics of North America - Small Animal Practice* **52**, 1359–1375 (2022).
9. Hong, A., Zampieri, R. A., Shaw, J. J., Floeter-Winter, L. M. & Laranjeira-Silva, M. F. One Health Approach to Leishmaniases: Understanding the Disease Dynamics through Diagnostic Tools. *Pathogens* **9**, 809 (2020).
10. Quinnell, R. J. & Courtenay, O. Transmission, reservoir hosts and control of zoonotic visceral leishmaniasis. *Parasitology* **136**, 1915–1934 (2009).
11. Franssen, S. U. *et al.* Global genome diversity of the Leishmania donovani complex. *Elife* **9**, e51243 (2020).
12. Negreira, G. H. *et al.* The adaptive roles of aneuploidy and polyclonality in Leishmania in response to environmental stress. *EMBO Rep.* **24**, e57413 (2023).
13. Dumetz, F. *et al.* Molecular Preadaptation to Antimony Resistance in Leishmania donovani on the Indian Subcontinent. *mSphere* **3**, e00548-17 (2018).
14. Smith, D. F., Peacock, C. S. & Cruz, A. K. Comparative genomics: From genotype to disease phenotype in the leishmaniases. *Int. J. Parasitol.* **37**, 1173 (2007).

15. Gommers-Ampt, J. H. *et al.* β-d-glucosyl-hydroxymethyluracil: A novel modified base present in the DNA of the parasitic protozoan T. brucei. *Cell* **75**, 1129–1136 (1993).

16. Grünebast, J., Lorenzen, S. & Clos, J. Genome-wide Quantification of Polycistronic Transcription in Leishmania major. *bioRxiv* 2023.11.23.568479 (2023) doi:10.1101/2023.11.23.568479.

17. Clayton, C. Regulation of gene expression in trypanosomatids: Living with polycistronic transcription. *Open Biol.* **9**, (2019).

18. Padilla-Mejía, N. E. *et al.* Gene organization and sequence analyses of transfer RNA genes in Trypanosomatid parasites. *BMC Genomics 2009 10:1* **10**, 232- (2009).

19. Mallik, S. *et al.* The evolutionary origin of host association and polycistronic transcription in trypanosomatids. *bioRxiv* 2025.10.27.684775 (2025) doi:10.1101/2025.10.27.684775.

20. Grünebast, J. & Clos, J. Leishmania: Responding to environmental signals and challenges without regulated transcription. *Comput. Struct. Biotechnol. J.* **18**, 4016 (2020).

21. Bussotti, G. *et al.* Genome instability drives epistatic adaptation in the human pathogen Leishmania. *Proc. Natl. Acad. Sci. U. S. A.* **118**, e2113744118 (2021).

22. Rogers, M. B. *et al.* Chromosome and gene copy number variation allow major structural change between species and strains of Leishmania. *Genome Res.* **21**, 2129 (2011).

23. Piel, L. *et al.* Experimental evolution links post-transcriptional regulation to Leishmania fitness gain. *PLoS Pathog.* **18**, e1010375 (2022).

24. Lahav, T. *et al.* Multiple levels of gene regulation mediate differentiation of the intracellular pathogen Leishmania. *The FASEB Journal* **25**, 515 (2011).

25. Piel, L. *et al.* Experimental evolution links post-transcriptional regulation to Leishmania fitness gain. *PLoS Pathog.* **18**, (2022).

26. dos Reis, M., Savva, R. & Wernisch, L. Solving the riddle of codon usage preferences: a test for translational selection. *Nucleic Acids Res.* **32**, 5036 (2004).

27. Späth, G. F. & Bussotti, G. GIP: an open-source computational pipeline for mapping genomic instability from protists to cancer cells. *Nucleic Acids Res.* **50**, e36–e36 (2022).

28. McDonald, J. R. *et al.* Localization of Epigenetic Markers in Leishmania Chromatin. *Pathogens* **11**, (2022).

29. Diotallevi, A. *et al.* Histone H3 K4 trimethylation occurs mainly at the origins of polycistronic transcription in the genome of Leishmania infantum promastigotes and intracellular amastigotes. *BMC Genomics* **26**, 167 (2025).

30. Grünebast, J., Lorenzen, S. & Clos, J. Genome-wide quantification of polycistronic transcription in Leishmania major. *mBio* **16**, (2025).

31. Harigaya, Y. & Parker, R. Analysis of the association between codon optimality and mRNA stability in Schizosaccharomyces pombe. *BMC Genomics 2016 17:1* **17**, 895- (2016).

32. Rak, R., Dahan, O. & Pilpel, Y. Repertoires of tRNAs: The Couplers of Genomics and Proteomics. *Annu. Rev. Cell Dev. Biol.* **12**, (2018).

33. Iben, J. R. & Maraia, R. J. tRNAomics: tRNA gene copy number variation and codon use provide bioinformatic evidence of a new anticodon:codon wobble pair in a eukaryote. *RNA* **18**, 1358 (2012).

34. Man, O. & Pilpel, Y. Differential translation efficiency of orthologous genes is involved in phenotypic divergence of yeast species. *Nature Genetics 2007 39:3* **39**, 415–421 (2007).

35. Shalem, O., Groisman, B., Choder, M., Dahan, O. & Pilpel, Y. Transcriptome Kinetics Is Governed by a Genome-Wide Coupling of mRNA Production and Degradation: A Role for RNA Pol II. *PLoS Genet.* **7**, e1002273 (2011).

Declaration of specific contributions

Figure	Data collected by	Data analyzed by	Notes (specify for individual panels if needed)
1	Me	Me	Panel C was fully created by Dr. Saurav Mallik
2	Me	Me	Panel A-F, the experiment was performed in Prof. Shulamit Michaeli's lab Panel B,C data was calculated in Gerald F. Späth's lab
3	Me and Dr. Saurav Mallik	Me and Dr. Saurav Mallik	
4	Me	Me	
5	Me	Me	
6	Me and Dr. Saurav Mallik	Me	
7	Me and Dr. Saurav Mallik	Me	

Acknowledgements

I wish to express my sincere gratitude to the many individuals whose support and expertise were vital to the completion of this research.

First and foremost, I am deeply grateful to Prof. Yitzhak Pilpel. His exceptional mentorship not only guided this thesis but also opened the door to the intellectual depth and beauty of scientific inquiry. I am indebted to him for fostering a rigorous scientific mindset, encouraging my curiosity, and instilling in me a lasting passion for exploring the world's mysteries.

My sincere thanks go to Dr. Orna Dahan, who served as a key bridge into this scientific community. I am grateful for her kindness, for her invaluable advice at every turn, and for her constant positive feedback, which greatly enriched my experience within the group.

I would also like to acknowledge Prof. Shulamit Micheli and Prof. Gerald F. Späth for their generous collaboration. Their introduction to the fascinating world of *Leishmania* and their commitment to scientific excellence were a source of inspiration throughout my work.

I am deeply thankful to Dr. Saurav Mallik for his patient guidance during my first steps in bioinformatics. He offered elegant solutions to daily technical challenges and set an example of dedication and ambition, teaching me the importance of striving for excellence in every detail.

To Dvir Dahary, thank you for the insightful discussions that helped shape my research perspective. His ability to balance broad conceptual thinking with meticulous attention to detail contributed greatly to my intellectual growth.

I also extend my warm appreciation to all members of the lab. The collaborative atmosphere, collective wisdom, and willingness to help made the research process truly rewarding.

Finally, I owe my most heartfelt thanks to my family: to my parents, for their unwavering support and for always encouraging me to aim higher; to my wife, who has been my constant anchor and whose patience allowed me to dedicate countless hours to this work; and to our children—you are my greatest source of energy and motivation.