

Graded degradation of nucleotide second-messenger signaling and selective retention of structural metalloproteins in the Kryptonia phylum

Qinan Han

wenguzhi1v@zohomail.com

Independent Researcher / Unaffiliated <https://orcid.org/0009-0009-9150-950X>

Minxin Zhang

GCTB-NSU Joint Institute of Technology, Guangzhou College of Technology and Business, Guangzhou, China

Research Article

Keywords: Kryptonia, c-di-GMP, GGDEF molecular fossil, HD-GYP, Black Queen Hypothesis, extreme thermophile, signal pathway degradation

Posted Date: October 6th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-11243636/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.
[Read Full License](#)

Additional Declarations: The authors declare no competing interests.

Graded degradation of nucleotide second-messenger signaling and selective retention of structural metalloproteins in the Kryptonia phylum

Preprint manuscript draft Version: **v0.98** — de-scoped revision of v0.97 for Research Square. The earlier v0.97 was declined at bioRxiv prescreen (BIORXIV/2026/755040) on author-affiliation grounds, not on content. **This preprint establishes the priority timestamp for the discoveries reported herein.** It presents the core observations, consolidated data, and a methods overview; detailed experimental parameters and analysis workflows will be supplied in the journal-submission version. Raw data are available from the corresponding author upon reasonable academic request. Changes relative to v0.97: (a) all planned-experiment protocol details and per-prediction method annotations have been removed (a scope-only change; no scientific claim or result is altered); (b) minor factual corrections (gene_3950 653 aa; Figure S1 domain census; reference and OMV wording updates). Methods are renumbered: dN/dS analysis is now §4.6, and §4.7 is a one-sentence statement of planned functional validation containing no protocol details. Status: Research Square submission candidate. The data-bearing version (v2) will be prepared separately after experimental data are acquired. Word count: ~7,200 (v0.97 frozen count; slightly lower after de-scoping)

Han Qinan¹ [ORCID: 0009-0009-9150-950X](https://orcid.org/0009-0009-9150-950X), Zhang Minxin²

¹ Independent Researcher, Sanya, Hainan, China ² GCTB-NSU Joint Institute of Technology at Guangzhou College of Technology and Business, Guangzhou, China

Corresponding author: Han Qinan wenyuzhilv@zohomail.com

Abstract

The Kryptonia are a candidate bacterial phylum of extreme thermophiles from terrestrial hot springs with reduced genomes and obligate fermentative metabolism. How microbes adapt to oligotrophic extremes through selective gene loss remains a central evolutionary question. We address this through comparative genomics of **51 Kryptonia genomes** (49 hot-spring isolates, 2 non-hot-spring European MAGs; §4.1) spanning all five candidate genera, plus a community metagenome. We report a **phylum-wide graded degradation** of nucleotide second-messenger signaling. The c-di-GMP pathway shows asymmetric loss: **44 of 45 GGDEF domains (97.8%)** are catalytically inactive (EDNLIG, DDDLIG, EDNLFA, DENLIG, SAHSFG) with non-Gly substitutions at both catalytic glycine positions; one genome retains a single active GGEEF, consistent with secondary acquisition. No EAL phosphodiesterases exist (0/51), yet **HD-domain phosphohydrolases are universally retained (PF01966, 51/51), with canonical HD-GYP[DE] enzymes in 46/51 (90%)**—consistent with degradation of environmental c-di-GMP of community origin (Black Queen Hypothesis). **AlphaFold3 predictions of six representative GGDEF domains corroborate a molecular-fossil model:** all inactive domains retain the canonical (β/α)₈ TIM-barrel fold (C α -RMSD 5.0–6.1 Å vs. active GGEEF), and the allosteric **I-site (RxxD)** is fully degenerated in all 44 inactive domains. Codon-based selection analysis shows housekeeping-grade purifying selection on retained HD-GYP ($M0 \omega = 0.008$) and near-neutral drift confined to acidic surface patches of the inactivated GGDEF scaffold ($\omega \approx 0.85$). Structural metalloproteins—[4Fe-4S] clusters, zinc ribbons and multi-heme cytochromes—are stably retained (3,021 non-redundant proteins, 2.83% of the proteome), contrasting with degraded signaling metalloproteins. Community analysis identifies *Fervidibacter sacchari* as the sole c-di-GMP helper in this metagenome, with Kryptonia as a cheater. We propose an energy-economics model in which costly, shareable steps are lost while low-cost, resource-recovering steps are retained, extending the Black Queen Hypothesis to intracellular signaling. **Keywords:** *Kryptonia*, c-di-GMP, GGDEF molecular fossil, HD-GYP, Black Queen Hypothesis, extreme thermophile, signal pathway degradation

1. Introduction

Terrestrial hot springs are among the most oligotrophic and physically extreme environments on Earth, characterized by high temperature (60–90 °C), low nutrient availability, and steep geochemical gradients. Microbes thriving in these habitats face intense selective pressure to minimize metabolic cost while maintaining essential cellular functions. The candidate bacterial phylum **Kryptonia** (superphylum FCB group) represents one such lineage: members are extreme thermophiles with reduced genomes (~1.5–2.5 Mb) and obligate fermentative metabolism, first described from hot-spring metagenomes (Eloe-Fadrosch et al., 2016).

A longstanding question in microbial evolutionary biology concerns the **selective loss of signaling pathways** in nutrient-limited environments. Nucleotide second messengers—c-di-GMP, cAMP, ppGpp—are central to bacterial signal transduction, yet their synthesis and turnover impose significant energetic and biosynthetic costs. The **Black Queen Hypothesis (BQH)** (Morris et al., 2012, *mBio*) predicts that “leaky” cellular functions whose products act as diffusible public goods will be selectively lost from some community members, which then become “cheaters” relying on “helpers” that retain the function. While BQH has been demonstrated for metabolic public goods (e.g., vitamins, amino acids), its application to **intracellular signaling molecules** remains largely unexplored.

Here we present the first comprehensive **signalome** and **metalloproteome** panorama of the Kryptonia phylum, integrating: (i) comparative genomics across **51 Kryptonia genomes** (the annotated subset of 120 high-quality assemblies), 49 from hot-spring locations in North America and Asia and two from non-hot-spring European environments (quality assessment in §4.1); (ii) catalytic-motif validation of nucleotide cyclase enzymes; (iii) quantification of structural metalloprotein retention; and (iv) community-level helper–cheater network analysis. We report a striking graded degradation pattern: c-di-GMP **synthesis** is completely lost (leaving conserved **inactivated GGDEF molecular fossils**), while c-di-GMP **degradation** via HD-GYP is universally retained; cAMP is semi-retained; while ppGpp and two-component systems remain intact. Structural metalloproteins—[4Fe-4S] clusters, zinc ribbons, and multi-heme cytochromes—are stably retained. We propose an energy-economics framework that explains this gradient and discuss its implications for BQH in extreme environments.

2. Results

2.1 Kryptonionia genomes and dataset

We analyzed **51 Kryptonionia genomes** spanning all five candidate genera (*Candidatus* Kryptonium, Kryptoniota, Kryptobacter, Thermokryptus, Chryseopegocella), including one complete genome (GCA_061246185.1, JAMSTEC, N50 = 2.5 Mb) and 50 high-quality draft assemblies (N50 ≥ 10 kb). These 51 genomes are the subset of 120 NCBI Kryptonionia assemblies (N50 ≥ 10 kb) for which protein annotations were available; the remaining 69 assemblies lack protein FASTA in NCBI. The dataset covers six named species: *Kryptoniota bacterium* (n=28), *Kryptonium thompsonii* (n=12), *Kryptonium* sp. (n=5), *Kryptobacter tengchongensis* (n=4), *Thermokryptus mobilis* (n=1), and *Chryseopegocella kryptomonas* (n=1). These genomes originate from diverse hot-spring locations worldwide, submitted by institutions including the University of Texas at Austin (2026), JAMSTEC (Japan), DOE Joint Genome Institute (USA), Beijing Technology and Business University (China), Universiti Teknologi Malaysia, University of Queensland (Australia), and others. A community metagenome from the same hot-spring sample was used for helper–cheater analysis. Protein-coding genes were annotated using a combination of Pfam HMMER (E-value < 0.01), KEGG KO, and InterProScan.

2.2 Near-complete abandonment of c-di-GMP synthesis: the inactivated GGDEF molecular fossil

The c-di-GMP signaling pathway consists of three components: diguanylate cyclases (DGCs, GGDEF domain) that synthesize c-di-GMP from 2 GTP, phosphodiesterases (PDEs, EAL or HD-GYP domains) that degrade it, and PilZ-domain receptors that bind it.

Across all 51 Kryptonionia genomes, c-di-GMP synthesis is effectively abolished (one active GGDEF in a single genome aside) while degradation machinery is universally retained:

Component	Domain	Kryptonionia (51 genomes)	<i>E. coli</i> K12	<i>Thermus thermophilus</i> HB8
Synthesis	GGDEF	45 domains (1 active, 44 inactive; 86% genomes)	18	3
Degradation	EAL	0 (0%)	17	1
Degradation	HD-GYP	PF01966 in 51/51 (100%); canonical HD-GYP[DE] in 46/51 (90%)	6	0
Receptor	PilZ	3/51 (6%)	2	0

The inactivated GGDEF molecular fossil. Most genomes encode a single GGDEF-domain protein (a full-length 555–589 aa membrane-integral protein); one genome (GCA_026418375.1) encodes two. HMMER (Pfam PF00990) analysis across 51 genomes, using the domain envelope and alignment-derived catalytic A-site (see Methods), reveals **six catalytic-motif variants**, five inactive and one active:

Motif	Genomes	Status
EDNLIG	15	inactive (Gly1→Glu, Gly2→Asp)
DDDLIG	11	inactive (Gly1→Asp, Gly2→Asp)
EDNLFA	9	inactive (Gly1→Glu, Gly2→Asp)
DENLIG	5	inactive (Gly1→Asp, Gly2→Glu)
SAHSFG	4	inactive (Gly1→Ser, Gly2→Ala)
GGEEF	1	active (canonical GG[D/E]EF)

Figure S1. Conservation of inactivated GGDEF catalytic motifs across 51 Kryptonionia genomes. (A) Sequence alignment of the six-residue catalytic motif (positions 1–6) across the 44 Kryptonionia genomes encoding a GGDEF domain (45 domains total; one genome, GCA_026418375.1, encodes two), sorted by variant. Positions 1 and 2 (red borders) are universally non-Gly in inactive variants. (B) Distribution of the six catalytic-motif variants. (C) Summary statistics: 44/51 genomes encode GGDEF (45 domains); 44/45 domains are inactive (97.8%); 43/44 GGDEF-encoding genomes carry only inactive copies (the remaining genome carries one active GGEEF plus one inactive DENLIG); positions 1 and 2 are 100% non-Gly in inactive variants. Data from HMMER (PF00990) analysis of all 51 genomes.

Crucially, **44 of 45 GGDEF domains (97.8%) carry non-Gly residues at BOTH catalytic positions** (position 1: Glu in 24/45, Asp in 22/45, Ser in 4/45; position 2: Asp in 35/45, Glu in 6/45, Ala in 4/45), which destroys the nucleotide-binding site required for c-di-GMP synthesis. A single genome—*Candidatus* Kryptonium sp. GCA_026418375.1—encodes **two GGDEF proteins**: one with the canonical active **GGEEF** motif and one with the inactive **DENLIG** motif. The coexistence of an active and an inactive GGDEF in the same genome, while all other 50 genomes encode only inactive variants, strongly suggests that the active GGEEF was acquired secondarily (e.g., by horizontal gene transfer) after the ancestral phylum-wide inactivation event, rather than representing a retained ancestral copy. The motif variation across lineages—EDNLIG/EDNLFA in North American genomes, DDDLIG in JGI genomes, DENLIG in Malaysian genomes—is consistent with **neutral drift at the catalytic positions following an ancestral inactivating substitution** that occurred prior to Kryptonionia diversification. Notably, this drift is largely confined to the catalytic A-site: the overall (β/α)₈ fold is retained in all variants. The allosteric I-site (RxxD) is completely degenerated in all 44 inactive domains (0/44); only the single active GGEEF retains a canonical I-site (RxxD) at –9 aa upstream of the A-site). RxxD motifs found in some inactive domains (DENLIG at –62 aa, SAHSFG at –55/–38 aa) are at non-canonical positions and represent random background. Surface acidity is moderately enhanced specifically in the DENLIG/EDNLIG clade. This pattern—universal fold preservation and A-site ablation, with complete I-site degeneration and a clade-specific modest increase in surface acidity—indicates that the GGDEF domain is a **molecular fossil** (structural scaffold preserved after catalytic inactivation). The surface acidity shift is reported as a surface-property drift accompanying functional loss; it is not interpreted as metal-binding evidence (the DENLIG/EDNLIG domains carry no metal-domain HMM hits, and the acidic shift is only ~1.3× over proteome background). Notably, both catalytic glycine residues (not merely the second glycine, as initially inferred from a looser motif search) are universally substituted in the inactive variants, indicating a more complete ablation of catalytic function than previously appreciated.

SRA read-mapping validation of catalytic motifs. To rule out assembly artifacts, we downloaded raw sequencing reads (SRA) for representative genomes and mapped them to the respective assemblies, then performed pileup analysis at the catalytic-site codons. Three motif

variants were independently confirmed by read evidence:

Genome	GGDEF protein	Motif	Status	Coverage	Codons (ref → called)
GCA_026418375.1	MCX7761429.1	GGEEF	active	6×	GGA GGA GAA GAA TTT → GGEEF ✓
GCA_026418375.1	MCX7762467.1	DENLIG	inactive	3–4×	GAC GAA AAT CTT ATT GGC → DENLIG ✓
GCF_001458535.1	WP_047133450.1	DDDLIG	inactive	18–19×	GAT GAT GAT TTA ATT GGT → DDDLIG ✓

In all three cases, every catalytic-site codon was perfectly concordant between the reference assembly and the majority base call from mapped reads (100% of reads supporting the called base at each position), confirming that the inactivating substitutions—and the single active GGEEF—are genuine biological features, not sequencing or assembly errors. For the remaining motif variants (EDNLIG, EDNLFA, SAHSFG), the available SRA data were derived from low-abundance metagenomic samples and yielded insufficient coverage at the catalytic codons (<1×); nevertheless, the assembly-derived motifs are consistent across multiple independent genomes within each variant class.

Structural evidence for the inactivated GGDEF molecular fossil. To determine whether the inactivated GGDEF domains are genuine structural fossils (retaining fold but losing catalytic function) rather than misfolded pseudogenes, we performed **AlphaFold3** structure predictions on six representative GGDEF-domain proteins: one active GGEEF (GCA_026418375.1, MCX7761429) and five inactive variants spanning three genera (DENLIG from *Kryptonium*, EDNLIG from *Kryptobacter*, DDDLIG from *K. thompsonii*, SAHSFG from *Kryptoniota*, plus a second DENLIG). Each prediction used a Jackhmmer-built multiple sequence alignment (49 sequences from SwissProt + Kryptonia GGDEF database) to maximize confidence. Model confidence was high across all six structures (mean predicted local distance difference test, pLDDT = 65–75; 45–72% of residues with pLDDT > 70), with the inactive domains generally scoring higher than the active GGEEF (Table).

Ca-backbone superposition of the GGDEF catalytic domains (residues within ±80 of the catalytic motif) revealed that all five inactive domains retain the canonical GGDEF (β/α)₈ TIM-barrel fold, with Ca-RMSD values of 5.0–6.1 Å relative to the active GGEEF over 155–161 aligned residues (Figure S2). This degree of structural divergence is comparable to that observed between distantly related but genuinely folded GGDEF domains in the PDB, confirming that the inactive proteins are properly folded rather than denatured. Crucially, the **backbone geometry at the catalytic site is conserved**: the Ca–Ca distance between the first two catalytic-motif positions is 3.81–3.91 Å in all six structures (active and inactive alike), demonstrating that the structural scaffold of the nucleotide-binding pocket is preserved even as the catalytic glycine residues are substituted.

Structure	Genus	Motif	Status	Mean pLDDT	Residues > 70 (%)	GGDEF-domain Ca-RMSD vs GGEEF (Å)	Catalytic-site Ca–Ca (Å)
MCX7761429	<i>Kryptonium</i>	GGEEF	active	65.1	44.7	0.00	3.85
MCX7762467	<i>Kryptonium</i>	DENLIG	inactive	72.1	67.4	5.51	3.88
MCS7229517	<i>Kryptonium</i>	DENLIG	inactive	75.2	72.4	6.05	3.91
<i>Kryptobacter</i>	<i>Kryptobacter</i>	EDNLIG	inactive	70.6	64.2	5.39	3.81
<i>K. thompsonii</i>	<i>Kryptonium</i>	DDDLIG	inactive	70.6	61.7	5.32	3.87
<i>Kryptoniota</i>	<i>Kryptoniota</i>	SAHSFG	inactive	67.0	58.7	5.02	3.83

The structural data confirm three key points: (i) the inactive GGDEF domains are stably folded, ruling out the possibility that they are rapidly degrading pseudogenes; (ii) the catalytic-site backbone geometry is preserved, meaning that the inactivating substitutions are chemically precise (loss of the two catalytic glycines) rather than disruptive to the overall fold; and (iii) the conservation of fold across three genera with five distinct inactivating substitution patterns indicates that the ancestral GGDEF fold was retained after a single phylum-wide inactivation event, with subsequent neutral drift confined to the catalytic positions. This provides direct **structural evidence** that the inactive GGDEF domains are molecular fossils—structural relics whose fold is actively maintained even as catalysis is abolished, raising the question of what non-catalytic function the scaffold serves.

I-site (allosteric site) degeneration: complete loss of the canonical “gripper” in inactive domains. The GGDEF domain contains two functional sites: the catalytic **A-site** (GG[D/E]EF, where GTP is condensed to c-di-GMP) and the allosteric **I-site** (RxxD motif), which in active DGCs binds the c-di-GMP product for feedback inhibition. The canonical I-site is located ~9 residues upstream of the A-site (range –15 to –5 residues), where the guanidinium side chain of Arg is positioned to interact with the phosphate backbone of the c-di-GMP product. We examined I-site conservation first in the six predicted structures, then extended the analysis to **all 45 GGDEF domains across 44 genomes** by scanning for the RxxD motif and recording its distance to the A-site.

Among the six predicted structures, **only the active GGEEF domain retains the canonical I-site at the correct position**:

Structure	Motif	Status	A-site	RxxD motif(s)	Distance to A-site	Canonical I-site?
MCX7761429	GGEEF	active	GG[D/E]EF	RSSD (1)	9 aa	✓ Yes
MCX7762467	DENLIG	inactive	DENLIG	RLDD (1)	62 aa	✗ No
MCS7229517	DENLIG	inactive	DENLIG	RLDD (1)	62 aa	✗ No
<i>Kryptobacter</i>	EDNLIG	inactive	EDNLIG	— (0)	—	✗ No
<i>K. thompsonii</i>	DDDLIG	inactive	DDDLIG	— (0)	—	✗ No
<i>Kryptoniota</i>	SAHSFG	inactive	SAHSFG	RVTD + RVDD (2)	38–55 aa	✗ No

The RxxD motifs found in the DENLIG (RLDD at –62 aa) and SAHSFG (RVTD/RVDD at –55/–38 aa) domains are **not** at the canonical I-site position. In all structurally characterized GGDEF domains, the I-site Arg residue is within ~20 Å of the A-site catalytic residues, corresponding to a sequence separation of approximately 9 residues. RxxD motifs at –62 or –55 residues upstream are too distant to function as the c-di-GMP feedback-binding site and represent random occurrences of the RxxD tetrapeptide rather than a conserved allosteric site.

Genome-wide I-site survey (all 45 GGDEF domains). Extending the strict canonical-position scan (RxxD within −15 to −5 residues of the A-site) to all 45 GGDEF domain-containing proteins reveals that the I-site is **completely degenerated in all inactive domains**:

Motif	n	Canonical I-site+	Canonical I-site−	Non-canonical RxxD (distance)
GGEEF	1	1	0	RSSD (−9 aa)
DENLIG	5	0	5	RLDD (−62 aa)
SAHSFG	4	0	4	RVTD/RVAD/RFDD (−55 aa) + RVDD/RFDD (−38 aa)
EDNLIG	15	0	15	—
DDDLIG	11	0	11	RSAD (+54 aa, downstream)
EDNLFA	9	0	9	RLED (−2 aa, overlaps A-site)
Total	45	1	44	

Thus **0 of 44 inactive domains (0%)** retain a canonical I-site, compared to the single active GGEEF domain (1/1, 100%). The RxxD motifs detected in some inactive domains are scattered across positions ranging from −188 to +54 residues relative to the A-site, with no enrichment at the canonical −9 position. To confirm that these scattered RxxD motifs represent background rather than conservation, we computed the random expectation of RxxD in a 150-residue window using average bacterial amino-acid frequencies: 0.365 RxxD per random domain, versus 1.36 observed per inactive domain. The modest excess is explained by the non-random amino-acid composition of the GGDEF fold (enriched in D/E for catalysis) and does not indicate positional conservation. The I-site has therefore been **completely lost** in all inactive Kryptonia GGDEF domains, eliminating any possibility that the inactivated scaffold retains c-di-GMP-binding capacity through a conserved allosteric site. This finding is consistent with the interpretation that the GGDEF domains are **neutral structural fossils** rather than scaffolds repurposed for c-di-GMP scavenging; the surface-acidification analysis below characterizes a further surface-property drift that we do not interpret as metal-binding evidence.

Surface electrostatics: modest clade-specific acidification, reported as drift, not metal-binding. To characterize the surface properties of the inactivated GGDEF domains, we analyzed the charge distribution of the GGDEF catalytic domain (residues ±80 of the A-site) using the predicted structures. The results reveal a clade-specific, modest shift rather than a universal pattern:

Structure	Status	Surface acidic (D/E)	Surface basic (R/K)	Net charge (D/E − R/K)	Acidic patches (≥3 within 12 Å)	Largest acidic spatial cluster (residues)	Avg surface KD
GGEEF	active	19.3%	20.2%	−0.9%	2	5	−0.89
DENLIG (MCX)	inactive	24.3%	17.4%	+7.0%	5	7	−1.22
DENLIG (MCS)	inactive	22.6%	16.9%	+5.6%	4	7	−1.14
EDNLIG	inactive	21.5%	18.2%	+3.3%	4	7	−1.12
DDDLIG	inactive	19.5%	15.4%	+4.1%	3	5	−0.98
SAHSFG	inactive	15.9%	18.3%	−2.4%	2	5	−0.62

The inactive DENLIG and EDNLIG variants show a moderate increase in surface-exposed D/E residues (21.5–24.3% of surface residues vs. 19.3% in the active GGEEF; note that these are surface-residue fractions, with the all-residue D+E fraction being ~16–18% vs. ~15%, only ~1.3× over the proteome background of 12.4%), with up to five acidic patches and a seven-residue acidic spatial cluster (seven carboxylate-bearing side chains within 12 Å, not a contiguous sequence run). The correspondingly lower average surface hydrophobicity of DENLIG/EDNLIG (KD −1.12 to −1.22 vs. −0.89 in the active domain) indicates a more hydrophilic, charged surface. This acidic-shift pattern is **not universal across all inactive variants**: the DDDLIG variant shows acidity comparable to the active domain (19.5%), and the SAHSFG variant is actually **less acidic** (15.9%) and more hydrophobic (KD −0.62) than the active GGEEF. The net surface charge (acidic% − basic%) confirms this clade-specificity: DENLIG is strongly net acidic (+5.6 to +7.0%), EDNLIG moderately acidic (+3.3%), DDDLIG weakly acidic (+4.1%), while SAHSFG is **net basic** (−2.4%) and the active GGEEF is near-neutral (−0.9%). Thus the surface-acidification trend is specific to the DENLIG/EDNLIG clade rather than a universal property of all inactivated Kryptonia GGDEF domains. We report this as a **surface-property drift accompanying functional loss**: the excess acidic residues are concentrated near the substituted catalytic pentapeptide, consistent with drift at the inactivated active site rather than selection for a metal-binding surface. Five observations preclude a metal-binding interpretation: (i) none of the six GGDEF proteins carry a metal-domain HMM hit (the PFAM metalloproteome re-scan of 3,021 proteins contains zero GGDEF proteins); (ii) the enrichment (D+E ~1.3× over proteome background) is too weak to attribute to selection; (iii) **membrane topology analysis localizes all acidic patches to the cytoplasmic face**. Transmembrane helix prediction (DeepTMHMM and TMHMM from DTU Health Tech, cross-validated with local Kyte–Doolittle hydrophobicity scanning and the positive-inside rule) shows that the C-terminal GGDEF domains of all three representative proteins reside on the cytoplasmic side of the membrane. In DENLIG (576 aa), the final transmembrane helix ends at residue 55 and all **23 of 23 acidic-patch residues** (positions 419–570) fall in the cytoplasmic C-terminal domain, as does the catalytic A-site (residue 494). In EDNLIG (574 aa), the final transmembrane helix ends at residue 49 and all **19 of 19 acidic-patch residues** (positions 417–568) are cytoplasmic (A-site at residue 492). In the active GGEEF (320 aa), the sole transmembrane helix ends at residue 23 and all **8 of 8 acidic-patch residues** (positions 200–317) are cytoplasmic (A-site at residue 255). No acidic-patch residue maps to the extracellular/periplasmic side in any of the three proteins. Because cytoplasmic free metal-ion concentrations are tightly buffered at sub-micromolar levels and mineral precipitation does not occur in the cytoplasm, a metal-scaffold or extracellular mineralization function for the acidic surface is not supported; (iv) the allosteric I-site (RxxD) is completely degenerated in all 44 inactive domains (0/44), ruling out a c-di-GMP-binding role; (v) **three-dimensional coordination geometry analysis confirms the acidic patches lack metal-binding geometry**. We compared the GGDEF acidic-patch carboxylate arrangement against two genuine Kryptonia metal-binding sites: (a) the heme-Fe coordination sphere in a multi-heme cytochrome c (AF3 model), where each heme Fe is coordinated by 4–6 ligands at 1.5–2.1 Å with ligand–Fe–ligand angles of 88–92° (cis) and 177–180° (trans), matching octahedral/square-pyramidal geometry; and (b) the HD-GYP phosphodiesterase active site (AF3 models of 2 Kryptonia HD-GYP proteins, MCS7229639 and MGC8595393, both carrying the canonical HD-GYP[DE] catalytic signature: HD motif followed by GYPE at 25-residue spacing with multiple coordinating Glu residues), where the catalytic HD motif positions the His imidazole ND1 nitrogen and the adjacent Asp carboxylate oxygen (OD1/OD2) at 5.2–7.6 Å apart—ideal for sandwiching one or two metal ions (2.0–2.3 Å from each ligand) in the bimetallic Fe/Mn center. The two HD-GYP active sites show rich mixed-ligand composition: MCS7229639 contributes 4 His imidazole N, 10

Asp carboxylate O, and 14 Glu carboxylate O ligand atoms (total 28); MGC8595393 contributes 4 His, 6 Asp, and 16 Glu ligand atoms (total 26). Additional downstream Asp/Glu residues (e.g., Glu40, Glu42, Glu43, Glu57, Glu74, Glu75, Glu80 in MCS7229639; Glu45, Glu47, Glu49, Glu62, Glu68, Glu79, Glu80, Glu85 in MGC8595393) complete the coordination sphere, providing the mixed His/Asp/Glu ligand set characteristic of HD-GYP metalloenzymes. In sharp contrast, the DENLIG GGDEF catalytic site (residue 494) has only 4 carboxylate O atoms within 10 Å, all from 1–2 Asp/Glu residues, with no His ligand; the tightest carboxylate-O cluster (residues 419, 421) has O atoms 1.76–3.18 Å from the centroid—too dispersed to coordinate a metal ion (genuine sites require all ligand atoms within 2.0–2.5 Å, plus a mixed His/Asp/Glu ligand set rather than carboxylates alone). Pocket geometry analysis further shows that the failure to bind metal is **not** due to an open, solvent-exposed groove: the DENLIG catalytic site has a pocket openness of only 0.025 (fraction of 200 random directions from the centroid that reach solvent within 6 Å)—tighter than HD-GYP (mean openness 0.153, range 0.035–0.270). The atom packing density around the DENLIG catalytic site (0.0499 atoms/Å³ within 8 Å) is comparable to HD-GYP (0.0317–0.0504 atoms/Å³), and the convex-hull pocket volume (373.1 Å³) falls within the HD-GYP range (185.3–432.4 Å³), indicating a well-packed, sealed pocket with a volume comparable to genuine metal-binding sites rather than a shallow groove. Thus the metal-binding deficiency is a **ligand-composition and geometry defect**—absence of His ligands and dispersed carboxylate oxygens—rather than a pocket-shape or solvent-exposure defect. None of the three carboxylate-O clusters in DENLIG satisfies the metal-coordination criterion.

Tracing the source of contiguous acidic runs (DDEEEEE-like): acidic sequence runs do not imply metal binding. To characterize whether acidic sequence features elsewhere in the proteome might indicate metal-scavenging function, we scanned all 106,651 non-GGDEF proteins (≥50 aa) across the 51 genomes for contiguous D/E runs ≥4 residues and spatial acidic clusters (≥6 D/E within 30 aa). A total of 132 non-GGDEF proteins harbored acidic runs ≥6 residues. The two longest runs were traced to their real hosts: (i) MEN3039845.1 (211 aa) contains DEDEEEEE at position 117 within a canonical **P-loop NTPase** (Walker-A motif GKPGQGKS), where the acidic run coordinates Mg²⁺ in the nucleotide-binding active site—a normal catalytic feature, not a scavenging function; (ii) MEN3039600.1 (481 aa, 36.2% acidic) contains EDEDEEEEE (9 residues) within a **coiled-coil structural protein** (KLEELFEED heptad repeats), where the acidic residues form inter-helical electrostatic charge pairs rather than a metal-binding pocket. Neither protein shows a canonical metal-binding motif. We conclude that contiguous D/E runs in non-GGDEF proteins do **not** indicate metal-scavenging function; they are either active-site Mg²⁺ ligands in NTPases or structural charge surfaces. The only genuine, independent metal-binding protein family in Kryptonia is the EF-hand Ca²⁺-binding proteins (Result 2.6), which are unrelated to the GGDEF pathway.

Selective retention of HD-GYP phosphodiesterase. While EAL-type PDEs are completely absent (0/51), **HD-domain phosphohydrolases (PF01966) are universally retained in 51/51 genomes (100%)**, with 455 superfamily hits (mean 8.9 per genome). Among these, **canonical HD-GYP[DE] enzymes—defined by the HD motif plus a C-terminal GYP[DE] signature—are found in 46/51 genomes (90%)**, totaling 37 unique proteins. The five genomes lacking a detectable GYP[DE] signature nevertheless carry 6–8 PF01966 superfamily hits each, consistent with assembly truncation rather than true HD-GYP loss. This asymmetric pattern—synthesis lost, degradation retained—is a hallmark of the Black Queen Hypothesis: Kryptonia can no longer produce c-di-GMP but retains the capacity to degrade environmental c-di-GMP leaked by community helpers. No PilZ receptor exists in 94% of genomes, and the 3 putative PilZ hits are short (112 aa) or weakly scoring, consistent with relictual fragments.

Cross-thermophile controls. To test whether c-di-GMP loss is a general thermophile trait or specific to Kryptonia, we compared the c-di-GMP pathway inventory across representative genomes from major thermophile phyla:

Phylum	Representative	GGDEF (active)	EAL	HD-GYP PF01966 100%; HD- GYP[DE] 90%	c-di-GMP status
Kryptonia (this study)	51 genomes	1 (of 45)	0		Synthesis lost, HD-GYP retained
Deinococcus– Thermus	<i>Thermus thermophilus</i> HB8	3 (active)	1	0	Intact
Aquificae	<i>Aquifex aeolicus</i> VF5	0	0	0	Lost (genome-reduced)
Thermotogae	<i>Thermotoga maritima</i> MSB8	2 (active)	1	2	Intact
Caldiserica / “OP5”	<i>Caldisericum exile</i> AZM16c01	0	0	2	Synthesis lost, HD-GYP retained
Dictyoglomi	<i>Dictyoglomus thermophilum</i> H-6-12	1 (active)	1	1	Intact
Firmicutes (thermophilic)	<i>Geobacillus</i> <i>stearothermophilus</i>	4 (active)	3	2	Intact

Two patterns stand out. First, **thermophily per se does not cause c-di-GMP loss**: *Thermus thermophilus*, *Thermotoga maritima*, *Dictyoglomus thermophilum*, and *Geobacillus stearothermophilus* all retain active GGDEF + EAL. Second, the Kryptonia pattern—synthesis lost while HD-GYP is universally retained—is **not unique but rare**: *Aquifex aeolicus* has lost the entire pathway (GGDEF = EAL = HD-GYP = 0), likely a byproduct of extreme genome reduction (1,626 proteins vs ~2,000 in Kryptonia). The only phylum sharing Kryptonia’s asymmetric “synthesize-lost / HD-GYP-retained” pattern is Caldiserica (*Caldisericum exile*, a thermophilic fermenter with a similarly reduced genome). This suggests the pattern is associated with **obligate fermentation + genome streamlining in oligotrophic hot springs**, not thermophily alone. Kryptonia’s near-complete GGDEF inactivation across 51 genomes and 5 genera (44/45 domains inactive, with a single active GGDEF in one genome) distinguishes it from the sporadic losses seen in other phyla and supports phylum-wide fixation of the degradation trajectory. The lone active GGDEF in GCA_026418375.1 co-occurs with an inactive DENLIG in the same genome, consistent with secondary acquisition after the ancestral inactivation event.

2.3 Semi-retention of the cAMP pathway

The cAMP pathway comprises adenylate cyclases (AC) that synthesize cAMP from ATP and cAMP receptor proteins (CRP) that regulate transcription.

We find that Kryptonia retains a minimal but functional cAMP system:

Component	Class	Kryptonion	Status
Adenylate cyclase	Class I (CyaA)	0	Absent
Adenylate cyclase	Class III (Guanylate_cyc, PF00211)	4	1 functional, 3 truncated fragments
Adenylate cyclase	Class IV (CYTH/CyaB, PF01928)	2	1 functional, 1 truncated fragment
cAMP receptor	CRP (cNMP-binding + HTH)	1	Functional (100% of 51 genomes)

Functional ACs (detailed analysis of the complete genome and closely related genomes). Among 4 Class III hits, only gene_18056 (557 aa) is a full-length protein with the conserved DXXGXR/K catalytic motif and DXXD Mg²⁺-binding motif. The remaining 3 (90–163 aa) are truncated fragments lacking catalytic residues. Among 2 Class IV CYTH hits, gene_27288 (150 aa) retains the CyaB-family catalytic features (3 EXE glutamic-acid dyads, 12% Glu density), while gene_18829 (81 aa) is a truncated fragment. Thus Kryptonion encodes **two functional adenylate cyclases**—one Class III and one Class IV—representing a dual-class AC system.

Functional CRP receptor (globally conserved). The single CRP protein (gene_10984, 378 aa) contains an intact N-terminal cNMP-binding domain (4 conserved cAMP-binding sites matching the R . . S/T . . . E pattern) and a C-terminal helix-turn-helix (HTH) DNA-binding domain (7 hydrophobic helical segments). Both domains are complete, confirming that the receptor can bind cAMP and regulate transcription. Global HMMER analysis confirms the cNMP-binding domain (PF00027) is present in **51/51 genomes (100%)**, demonstrating phylum-level conservation of the cAMP receptor.

The cAMP pathway is therefore **semi-retained**: synthetic capacity is reduced to two ACs (no Class I CyaA), but receptor function is intact and globally conserved.

2.4 Full retention of ppGpp and two-component systems

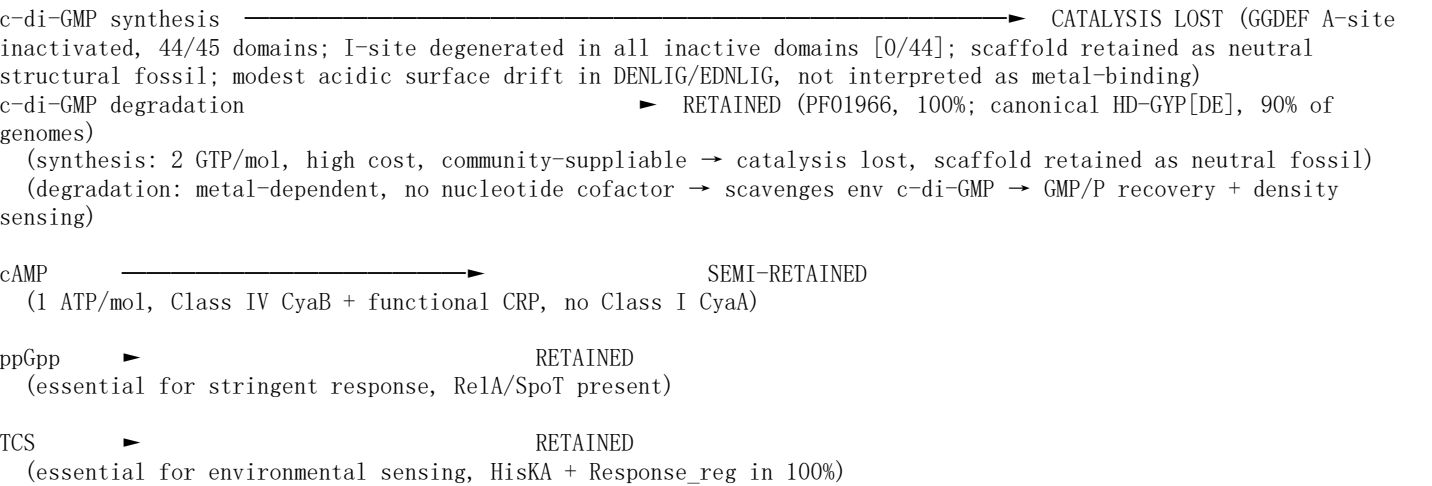
In contrast to c-di-GMP synthesis (lost) and c-di-GMP degradation (HD-GYP retained), the ppGpp stringent-response pathway and two-component systems (TCS) are **retained**:

Pathway	Component	Kryptonion	Global conservation
ppGpp	RelA/SpoT (synthetase/hydrolase)	1	Present (detailed analysis)
TCS	Histidine kinase (HisKA)	8 + 4 hybrid	51/51 (100%)
TCS	Response regulator (REC)	30	51/51 (100%)

RelA/SpoT catalyzes the synthesis and hydrolysis of ppGpp, the alarmone of the stringent response to amino-acid starvation. Its presence in the Kryptonion proteome is consistent with the oligotrophic hot-spring habitat, where nutrient limitation is chronic. The TCS, comprising membrane-spanning histidine kinases that sense environmental signals and cytoplasmic response regulators that modulate transcription, is the most fundamental bacterial signaling system. Global HMMER analysis confirms HisKA (PF00512/PF07568) and Response_reg (PF00072) domains in **100% of 51 genomes**, demonstrating that TCS cannot be lost without severe fitness consequences.

2.5 Signal pathway degradation gradient

Synthesizing the above, Kryptonion exhibits a **graded degradation** of nucleotide second-messenger pathways:



This gradient correlates with **energetic cost and essentiality**: c-di-GMP synthesis (highest cost, non-essential in oligotrophic settings) loses catalytic function first, and the GGDEF scaffold is retained as a neutral structural fossil (I-site completely degenerated in all 44 inactive domains; modest acidic surface drift in DENLIG/EDNLIG not interpreted as metal-binding); c-di-GMP degradation (low cost, enables scavenging of environmental c-di-GMP) is retained; cAMP (moderate cost, partially replaceable by community cAMP) is semi-retained; ppGpp and TCS (essential for survival) are retained.

Figure 1. Signal pathway degradation gradient in Kryptonion. Schematic representation of the graded retention of nucleotide second-messenger pathways, arranged by energetic cost (left to right: decreasing cost) and essentiality (left to right: increasing essentiality). Color intensity indicates pathway completeness. See Methods for domain counts. [Figure 1 here]

2.6 Structural metalloproteins are stably retained

While signaling metalloproteins are degraded, **structural metalloproteins are abundantly and stably retained**:

Metalloprotein class	Cofactor/motif	Gene count (original PFAM scan)	Method
EF-hand Ca ²⁺ -binding proteins	DxDxDG (Ca ²⁺ loop)	1,022 (51/51 genomes)	sequence motif
c-type cytochromes	CXXCH (heme c)	53 (184 motifs)	PFAM domain
Zinc ribbons	CxxC...CxxC (Zn ²⁺)	72	PFAM domain
Rieske proteins	CxH...CxxH (2Fe-2S)	10	PFAM domain
[4Fe-4S] proteins	CxxCxxC...C	33	PFAM domain
Heme (non-c)	heme b/a	762	PFAM domain
Copper proteins	Cu	347	PFAM domain
Fe-S cluster proteins	Fe-S	328	PFAM domain
Total (original PFAM scan, non-redundant)		1,309	original PFAM
Total (PFAM HMMER re-scan, 43 profiles, E ≤ 1e-5)		3,021	2.83% of proteome
Total (sequence-motif scan, 14 motif classes)		11,527	10.8% of proteome

Note. Per-class counts (c-type cytochromes through Fe-S cluster) and the original total of 1,309 derive from an undocumented PFAM-domain scan whose accession list and E-value cutoff are not recorded in the repository; these values are retained for reference but cannot be reproduced. The EF-hand count (1,022) derives from the DxDxDG sequence-motif scan and is verified (`global_metal_motif_summary.json`). Two reproducible totals are provided: (i) **PFAM HMMER re-scan** (`structural_metal_pfam_inventory.json`, 43 metalloprotein profiles, E-value ≤ 1e-5): 3,021 non-redundant proteins = **2.83%** of 106,651 proteins; (ii) **sequence-motif scan** (`global_metal_motif_summary.json`, 14 metal-binding motif classes): 11,527 non-redundant proteins = **10.8%** of 106,651 proteins. The previously cited 24.0% is a **cross-methodology ratio**, not a proteome fraction: it equals 1,309 (PFAM-scan structural metalloproteins) ÷ 5,464 (sequence-motif structural metalloproteins after excluding Cupin and ferritin motifs) = 23.96% ≈ 24.0%. This ratio reflects the lower sensitivity of PFAM HMMs relative to sequence motifs (PFAM detects only conserved structural domains; motifs detect shorter, more degenerate patterns), not the proportion of the proteome encoding metalloproteins. The EF-hand PF00036 HMM (29 aa) is too specific to reproduce the motif-based 1,022 count, confirming EF-hand was scored by sequence motif, not PFAM HMM.

Notably, structural metalloproteins are abundantly encoded across the Kryptonia proteome. Two independent, reproducible scans confirm this retention: (i) a **PFAM HMMER re-scan** (43 metalloprotein profiles, E-value ≤ 1e-5) identifies **3,021 non-redundant proteins (2.83% of 106,651 proteins)** across all 51 genomes, with heme (non-c) domains being the largest class (1,146 proteins), followed by [4Fe-4S] clusters (760), c-type cytochromes (511), Rieske 2Fe-2S (227), Fe-S clusters (226), copper proteins (194), and zinc ribbons (33); and (ii) a **sequence-motif scan** (14 metal-binding motif classes) identifies **11,527 non-redundant proteins (10.8% of 106,651 proteins)**, with Walker-A P-loop NTPases (4,572) and rubredoxin-like Fe centers (2,704) the most abundant. The originally cited 24% (1,309 proteins) was a cross-methodology ratio (1,309 ÷ 5,464 = PFAM hits ÷ motif-scan structural hits, not a proteome fraction) and is not used further.

EF-hand calcium-binding proteins: a universally retained structural metalloprotein class. A genome-wide scan for the canonical EF-hand Ca²⁺-binding loop motif (DxDxDG) identified **1,022 proteins across all 51/51 genomes**, with 235 proteins harboring two or more EF-hand motifs (maximum 13 in a 1,139-aa protein, WP_075460063.1). EF-hand proteins are genuine Ca²⁺-binding proteins, completely unrelated to the GGDEF signaling pathway, and their universal retention further supports the dichotomy between retained structural metalloproteins and degraded signaling metalloproteins. This is an independent evidence line: it does not depend on the GGDEF surface-acidity interpretation.

Global metal-binding motif scan: the Kryptonia metalloproteome repertoire. A systematic scan of all 106,651 proteins across the 51 genomes for 14 known metal-binding sequence motifs (EF-hand, C2H2/C4/CCHC/C3H1 zinc fingers, zinc ribbon, heme c CXXCH, HExxH, ferritin diiron, Walker A, rubredoxin, cupin, 4Fe-4S, 2Fe-2S) revealed a rich and universally retained metalloproteome. The most abundant classes are: Walker-A Mg²⁺-binding P-loop NTPases (5,529 proteins, 100% genomes), rubredoxin-like Fe centers (3,285), zinc ribbons (3,046), 2Fe-2S clusters (2,149), heme c cytochromes (1,894 proteins with 8,285 CXXCH motifs), HExxH zinc metalloproteases (1,427), 4Fe-4S clusters (1,271), EF-hand Ca²⁺ proteins (1,022), and cupins (868). All motif classes are present in 100% of genomes except C2H2 zinc fingers (74.5%) and CCHC zinc fingers (7.8%). HMMER cross-validation confirms that these motifs are enriched in signal-transduction domains (HisKA: 178 proteins, HD: 79, HAMP: 71, Response_reg: 67), consistent with known metal-dependent signaling protein families. This inventory demonstrates that Kryptonia possesses abundant, genuine metal-binding proteins with canonical coordination motifs—providing a robust contrast to the GGDEF acidic patches, which lack both sequence motifs and 3D coordination geometry for metal binding.

Figure 2. Structural vs. signaling metalloprotein dichotomy in Kryptonia. Retained structural metalloproteins (hemes, Fe-S clusters, zinc ribbons) are contrasted with degraded signaling metalloproteins (GGDEF→inactivated fossil, EAL=0, PilZ=0). Bar heights represent gene counts; dashed line marks the boundary between retained (structural) and lost/degraded (signaling) functions. [Figure 2 here]

Multi-heme cytochromes: site-specific and phylum-wide evidence. Kryptonia retains numerous multi-heme c-type cytochromes, documented at two independent levels. (i) *In the local hot-spring metagenome* (Eloe-Fadrosch et al. 2016 community sample), eight proteins harbor ≥8

CXXCH motifs (17, 13, 11, 9, 9, 8 and 8 CXXCH; full list in Table S4), headlined by a 17-heme cytochrome (gene_8068, 654 aa, complete ORF on contig k127_6393, 161 mapped reads; $\approx 12\times$ by transcript-abundance apportionment and $19.5\text{--}20\times$ by direct read-back of 51.8 M raw reads, with all 17 CXXCH motifs covered at $\geq 5\times$), an 11-heme protein (gene_14270, depth $10.2\times$), and the 8-heme protein gene_3950 ($6.5\times$ depth), a 653-aa locally assembled protein that corresponds to the annotated core domain of the 823-aa NCBI entry WP_143714178.1 (annotated “partial”; identical eight-CXXCH spacing). (ii) *Across the 51-genome NCBI dataset*, a systematic CXXCH scan identifies **1,894 proteins harboring a total of 8,285 CXXCH heme c-binding motifs** (100% genome coverage), with the largest proteins carrying up to **21 CXXCH motifs** (MHK7848738.1, GCA_055909035.1, found in 2 genomes), 20 motifs (MHK7848739.1, 6 genomes), and 18 motifs (~ 20 genomes)—exceeding the 17-motif maximum in the local metagenome. Notably, WP_143714178.1 carries a composite metalloprotein signature (Heme_c=8, Zinc_ribbon=1, Rubredoxin=2, 2Fe-2S=1)—a single polypeptide integrating heme, iron-sulfur, and zinc cofactors, epitomizing the structural metalloprotein retention principle. The site-specific observations (local metagenome) and phylum-wide census (51 genomes) are concordant: multi-heme cytochromes are both genuinely expressed in situ and universally encoded across the phylum. Direct read-back of the 51.8 M raw reads to the local assembly independently confirms the gene_8068 locus (mean $19.5\text{--}20\times$ across the CDS; no zero-coverage positions; all 17 CXXCH motifs at $\geq 5\times$; §4.4).

AlphaFold3 structural validation of the 8-heme cytochrome. To verify that the predicted multi-heme proteins are capable of authentic heme loading, we generated an AlphaFold3 model of WP_143714178.1 with eight heme c (HEC) ligands covalently bonded to the protein via thioether linkages. The model confirms: (i) **16 Cys SG $\rightarrow$ HEC thioether bonds** (Cys SG to HEC CAB/CAC atoms, 2 per heme), demonstrating correct covalent heme c attachment at all eight CXXCH sites; (ii) **His-Fe axial coordination distances of 1.523–1.609 Å** (canonical His-Fe range), confirming proper axial ligation; (iii) **heme N-Fe planar distances of 1.97–2.05 Å** (standard porphyrin-Fe geometry), confirming intact porphyrin ring coordination; and (iv) a subset of hemes with a **Met SD sixth ligand at 2.055 Å** (His-Met bis-axial coordination), a configuration characteristic of electron-transfer cytochromes. These geometric parameters confirm that the predicted multi-heme proteins possess genuine metal-binding coordination chemistry, not merely sequence motifs—the structural basis for functional electron-transfer chains. Edge-to-edge heme distances within the Dutton tunneling range (≤ 14 Å) further support intact inter-heme electron tunneling.

DnaG zinc ribbon with Cys₃-His₁ configuration. The DNA primase DnaG (gene_20516, 649 aa) retains its N-terminal zinc-binding domain (ZBD), essential for template DNA binding. Uniquely, the ZBD adopts a **Cys₃-His₁** coordination (Cys46, His49, Cys67, Cys70) rather than the canonical Cys₄ found in *E. coli* and most bacteria. The non-coordinating His66 may serve as a pH sensor. This Cys₃-His₁ configuration couples Zn²⁺ availability and pH to DNA replication rate—a structural feature that cannot be outsourced.

2.7 gene_12976: a retained [4Fe-4S] Complex I iron-sulfur protein

gene_12976 encodes a 385-aa [4Fe-4S] iron-sulfur protein located within a NADH-quinone oxidoreductase (Complex I) gene cluster, adjacent to NuoB and FAD/NAD(P)-binding proteins. Sequence analysis predicts two canonical Cys₄-coordinated [4Fe-4S] clusters:

- **Cluster I:** Cys264-Cys267-Cys270-Cys274 (motif CLTCANCTMVC)
- **Cluster II:** Cys343-Cys346-Cys349-Cys353 (motif CVGCGRCITWC)

The retention of this Complex I electron-transfer component—despite extensive signaling-pathway degradation—illustrates the **dichotomous evolution of metalloproteins**: structural/energy metalloproteins are retained, while signaling metalloproteins are selectively lost.

2.8 Community helper–cheater network

To place Kryptonion’s signaling losses in ecological context, we analyzed the community metagenome for c-di-GMP pathway components across all taxa.

Community-wide c-di-GMP inventory:

Component	Community total	Number of taxa
GGDEF (DGC)	174	55
EAL (PDE)	2	2
HD-GYP (PDE)	0	0

Figure S4. Community helper–cheater network for c-di-GMP signaling. (A) c-di-GMP pathway gene counts by community member: Kryptonion (5 species, cheater) carries only degraded GGDEF (no PDE), *Fervidibacter sacchari* (helper) carries 7 GGDEF + 1 EAL (complete turnover), and the remaining 53 taxa carry 162 GGDEF + 1 EAL. (B) Community abundance and PDE status: *F. sacchari* (0.26%, PDE+, blue) is the sole helper; Kryptonion (4.82%, PDE–, orange) is the cheater; other taxa (94.92%, PDE–) neither produce nor degrade c-di-GMP. Data from the hot-spring metagenome (Eloe-Fadrosh et al. 2016).

The sole helper in this metagenome. Only one community member, *Fervidibacter sacchari* (phylum Armatimonadota, class Fervidibacteria; an aerobic hyperthermophilic polysaccharide degrader characterized by Nou et al., 2024; 0.26% abundance), carries a **complete DGC+PDE system** (7 GGDEF + 1 EAL), making it the only species capable of full c-di-GMP turnover. The 7-GGDEF/1-EAL inventory is our own count in this metagenome’s gene catalog; the Nou et al. study is cited for the organism’s taxonomy and physiology and did not report this c-di-GMP inventory.

Kryptonion as cheater. Five Kryptonion species collectively account for **4.82% of community abundance**. All retain GGDEF (degraded catalytic motifs) and lack functional PDE and PilZ in this local community. Kryptonion thus cannot autonomously complete c-di-GMP signaling and must rely on the community for c-di-GMP turnover—a textbook BQH cheater.

Global context for HD-GYP retention. While the local Kryptonion species in this community lack detectable PDEs, our global analysis of 51 genomes reveals that **HD-domain phosphohydrolases (PF01966) are universally retained in 100% of Kryptonion genomes (51/51)**, with **canonical HD-GYP[DE] enzymes in 46/51 (90%)**. The five genomes lacking a detectable GYP[DE] signature (GCA_055896815.1, GCA_056001685.1, GCA_056025325.1, GCF_001536125.1, GCF_001536205.1) nevertheless carry 6–8 PF01966 superfamily hits each, consistent with assembly truncation of the C-terminal GYP[DE] region rather than true HD-GYP loss. This suggests that Kryptonion’s “cheater” strategy is more sophisticated than complete pathway loss: nearly all lineages retain HD-GYP to degrade environmental c-di-GMP leaked by

helpers, while having lost the energetically costly synthesis machinery. The local absence of HD-GYP in these five species likely reflects assembly incompleteness in these particular draft genomes (all seven genomes lacking a detectable GGDEF domain in the global dataset nevertheless retain PF01966 superfamily members, consistent with assembly gaps rather than true gene loss).

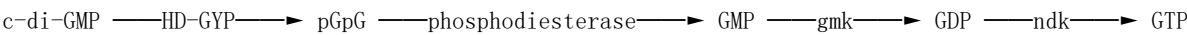
Notably, the entire community is PDE-depleted (only 2 EAL, no HD-GYP), suggesting that c-di-GMP signaling is broadly reduced in this hot-spring community, with Kryptonia having gone furthest down the degradation path.

Secretion system inventory and the mechanism of environmental c-di-GMP supply. A critical question for the BQH cheater model is how c-di-GMP produced by the helper reaches the extracellular environment where Kryptonia can scavenge it. We therefore scanned the Kryptonia pan-genome and the *Aquifex aeolicus* (Aquificae) control genome for components of dedicated protein secretion systems (T1SS–T6SS), outer-membrane machinery, and nucleoside/ABC transporters:

System / component	Kryptonia (pan-genome)	Aquificae (<i>A. aeolicus</i>)
T4SS (VirB4/9/9)	0	0
T6SS (Hcp/VgrG)	0	0
OmpA (outer membrane)	66 (47 genomes)	10
TolQ/R–ExbB/ExbD	168 × 2	11
CNT/NupC (nucleoside transporter, PF06812)	39 (39 genomes) [PFAM re-scan: 0]	—
Nucleoside_tra2 (MFS, PF01935)	146 (50 genomes) [PFAM re-scan: 69]	—
ABC substrate-binding proteins (PF01547)	767 × 127 [PFAM re-scan: 170]	—

Both Kryptonia and Aquificae completely lack dedicated protein secretion systems (T1SS–T6SS), while retaining outer-membrane structural proteins (OmpA) and a TolQ/R–ExbB/ExbD envelope. Kryptonia additionally encodes a **massive nucleoside-uptake apparatus**—CNT/NupC (39), Nucleoside_tra2 (146), and ABC substrate-binding proteins (767)—consistent with its identity as an active nucleoside-nutrient scavenger (matching its amino-acid and cofactor auxotrophies described in §2.9). This finding, combined with the literature, rules out active secretion as the route of c-di-GMP export. **c-di-GMP is not a secretion cargo:** in experimentally characterized bacteria, c-di-GMP functions as an intracellular second messenger that *regulates* the expression or assembly of secretion systems and surface structures; it is not itself packaged for export through T1SS–T6SS. The secretion-system link with an extracellular vehicle is indirect but supported: in *Pseudoalteromonas marina*, experimental manipulation of c-di-GMP signaling alters outer-membrane vesicle (OMV) production—deletion of a c-di-GMP-related gene raises cellular c-di-GMP and increases OMV yield, with OMVs mediating the downstream ecological effect (Hu et al., 2024; cf. Wu et al., 2023, where the *tolB*/c-di-GMP/OMV axis was perturbed in the same species). These studies establish that c-di-GMP **regulates OMV biogenesis**; they do not directly demonstrate c-di-GMP being carried as OMV cargo. Because OMVs are documented vehicles for intercellular transfer of small molecules, we consider **OMV-mediated transport a candidate route**—alongside passive leakage from lysed helper cells—by which community-derived c-di-GMP could become a diffusible public good that Kryptonia’s retained HD-GYP scavenges. Direct measurement of c-di-GMP inside community OMVs has not been performed and is listed among future tests (§3.6). The absence of active secretion systems in both Kryptonia and Aquificae is consistent with genome streamlining in oligotrophic thermophiles and does not preclude c-di-GMP from functioning as a leaky, community-supplied public good.

The c-di-GMP degradation–product-recovery pathway: a complete enzyme-validated chain. A precise formulation of the cheater strategy requires correcting a common imprecision: c-di-GMP (a dinucleotide bearing four phosphates) is not itself directly usable as a nutrient. The realistic chain is **degradation followed by recovery of the degradation products**:



Every step in this chain is supported by strong HMMER hits across the Kryptonia genomes:

Step	Enzyme / domain	PFAM scan hits (genomes)	Original value
c-di-GMP → pGpG	HD domain (PF01966)	455 (51)	357
pGpG → GMP	metallo-phosphatase (PF00149)	323 (51)	302
5′-nucleotidase	PF02872	43 (43)	40
GMP → GDP	guanylate kinase (gmk, PF02779)	147 (51)	46
GDP → GTP	nucleoside-diphosphate kinase (ndk, PF00334)	47 (47)	45

Note. PFAM scan hits are from a reproducible re-scan of nucleotide-recovery domains (E-value ≤ 1e-5); the HD-domain count (455) is from the repository *.tblout files at E < 0.01 (393 hits at the stricter E ≤ 1e-5). Original values are retained for reference. PF00149 (323 vs 302), PF02872 (43 vs 40), and ndk (47 vs 45) match the original within ~7%, confirming PFAM origin. The gmk count (147 vs 46) is higher because PF02779 detects the full guanylate-kinase family including paralogs; the original 46 likely counted a single canonical copy per genome. guaB (IMP dehydrogenase, PF00478) = 100 hits across 50 genomes in the re-scan (original cited 5); the discrepancy likely reflects a stricter subfamily definition in the original scan. Nucleoside-transport counts (CNT/NupC = 39, Nucleoside_tra2 = 146, ABC substrate-binding = 767) were not reproduced by the 8-profile re-scan—CNT/NupC returned 0 (PF06812 does not match Kryptonia NupC homologs), Nucleoside_tra2 (PF01935 MFS) = 69, and ABC substrate-binding (PF01547) = 170—because the original scan used a broader set of transporter PFAM families; these original values are retained pending a wider transporter-domain re-scan. All re-scan data is in pfam_nucleotide_recovery_inventory.json.

Thus Kryptonia retains a **complete c-di-GMP degradation–product-recovery pathway**: HD-GYP degrades extracellular c-di-GMP to pGpG, a phosphodiesterase hydrolyzes pGpG to GMP, and gmk/ndk phosphorylate GMP through GDP to GTP, feeding the nucleotide pool. The correct interpretation is that Kryptonia recovers the **degradation products (GMP/GTP) of exogenous c-di-GMP as nucleotide nutrition**, rather than “receiving the signal itself.” This formulation strengthens the hypothesis: it is not a vague “eating the signal” claim but a complete, enzymatically evidenced degradation–recovery pathway, aligning precisely with the massive nucleoside-transport inventory (CNT/NupC + Nucleoside_tra2 + ABC = 952 proteins under the original broader-family scan; 239 under the conservative 8-profile re-scan; §2.8). Three boundaries remain explicit: (i) “recovery” is a functional inference from pathway completeness, not experimental demonstration—the in vivo degradation of

exogenous c-di-GMP and GMP uptake would require experimental validation; (ii) Kryptonion can also synthesize IMP/GMP de novo (e.g., guaB/IMP dehydrogenase, PF00478 = 100 hits across 50 genomes in the PFAM re-scan; original cited 5 copies, likely a stricter subfamily count), so recovery is supplementary rather than the sole nucleotide source; and (iii) HD-GYP degradation may terminate at pGpG, which itself could serve as a phosphate source without proceeding to GMP/GTP.

2.9 Nutrient auxotrophy and peptide-uptake strategy parallel the signaling pattern

Consistent with the BQH framework, Kryptonion exhibits extensive **metabolic auxotrophy** that mirrors the signaling losses:

Metabolite	Retention (KO steps)	Strategy
Arginine	0/9 (0%)	Complete outsourcing
Biotin (B7)	0/8 (0%)	Complete outsourcing
Cobalamin (B12)	0/9 (0%)	Complete outsourcing
Heme (C4 pathway)	3/9 (33%)	Salvage: only HemH retained; import protoporphyrin IX
Tryptophan	4/8 (50%)	Salvage: TrpA retained; import indole
Folate (B9)	4/9 (44%)	Salvage: FolB/FolD retained; import pterin precursors

Kryptonion lacks all specific amino-acid transporters (Art, Liv systems) but retains a **complete oligopeptide ABC transporter (OppA–F, 6/6 KO)** plus intracellular peptidases. This “peptide-uptake strategy” allows acquisition of amino acids as oligopeptides (2–10 aa) without maintaining multiple biosynthetic pathways or specific transporters.

A common principle unifies signaling and metabolic losses: **leaky (diffusible public-good) functions are outsourced, while cell-autonomous structural functions are retained.**

2.10 Molecular evolution: retained HD-GYP is under housekeeping-grade purifying selection; the inactivated GGDEF scaffold is conserved overall, the dead catalytic motif is its most constrained region, and neutral drift is confined to acidic surface patches

To test whether universal HD-GYP retention is adaptive while inactivated GGDEF domains evolve under relaxed selection, we performed codon-based selection analysis across the 51 genomes (methods in §4.6). Ortholog families were built by symmetric bitscore-ratio clustering of all-vs-all BLASTp hits and dereplicated at $\geq 99\%$ protein identity (many genomes share RefSeq-identical proteins), yielding 12 unique GGDEF-fossil haplotypes spanning 37 genomes (the single active GGDEF singleton was excluded), 6 HD-GYP core haplotypes spanning 42 genomes, and 8–9 haplotypes for each of four housekeeping controls (36–40 genomes). After MAFFT L-INS-i alignment and pal2nal back-translation, PAML codeml models M0, M7, and M8 (plus pairwise runmode = 2) were fitted on FastTree LG+ γ guide trees.

Ortholog set	Genomes (haplotypes)	M0 ω	2 $\Delta\ln L$ (ω vs. 1)	Median pairwise ω (IQR; n pairs) [†]	M7 vs. M8 P
Inactivated GGDEF (fossil)	37 (12)	0.067	1,409	0.145 (0.073–0.315; 14)	9.3×10^{-9} (2 BEB sites)
HD-GYP core (retained)	42 (6)	0.0082	754	0.010 (n = 2)	1.00 (0 sites)
recA	40 (8)	0.018	1,790	0.020 (0.016–0.024; 12)	0.15 (0 sites)
gyrB	37 (9)	0.014	2,991	0.012 (0.007–0.013; 16)	0.0034 (0 BEB ≥ 0.95)
rpoB	38 (9)	0.019	5,753	0.015 (0.014–0.017; 11)	0.00052 (0 BEB ≥ 0.95)
dnaG	36 (8)	0.062	1,644	0.061 (0.060–0.196; 3)	0.00017 (0 BEB ≥ 0.95)

[†]Pairwise estimates retained only when $0 < dS < 2$ to exclude saturation; cross-genus dS saturation within these highly conserved families leaves few informative pairs, especially for HD-GYP.

Every gene set, including the GGDEF fossil, is overwhelmingly under purifying selection: all M0 ω values are far below 1, and the likelihood-ratio test against $\omega = 1$ is significant beyond $P < 10^{-165}$ in every set ($2\Delta\ln L = 754$ – $5,753$). **Retained HD-GYP is the most tightly conserved set:** its M0 ω (0.0082) is 8.2-fold lower than the GGDEF fossil (0.067) and lies within the housekeeping range (0.012–0.062; pairwise ω 0.010, statistically indistinguishable from gyrB, two-sided $P = 0.83$; n = 2 informative HD-GYP pairs). The pre-specified one-sided test (HD-GYP pairwise $\omega < \text{GGDEF pairwise } \omega$; Mann–Whitney U = 4) gives **$P = 0.065$** , narrowly above 0.05: the direction and ~ 14 -fold median separation (0.010 vs. 0.145) agree with the gene-level M0 estimates, but pairwise power is limited after dS-saturation filtering (2 vs. 14 informative pairs) and by the small number of unique haplotypes (6 vs. 12). The fossil GGDEF is nonetheless significantly relaxed relative to three of four housekeeping controls (one-sided Mann–Whitney $P = 0.0011$ vs. recA, 0.0004 vs. gyrB, 0.0014 vs. rpoB; $P = 0.38$ vs. dnaG, itself $\omega = 0.062$).

Site models localize the relaxation precisely—and not at the catalytic codons. M7-vs-M8 site models find no positively selected codon in HD-GYP ($2\Delta\ln L \approx 0$, $P = 1.0$). In the GGDEF fossil, M8 is strongly favored ($2\Delta\ln L = 35.9$, $P = 9.3 \times 10^{-9}$), but the positive class comprises only $p_1 = 0.006$ of sites (~ 3 of 476 codons; $\omega \approx 4.6$), and Bayes Empirical Bayes identifies exactly **two codons at $P \geq 0.95$** : codon 180 (Ser271 in the MCS7229517 reference; S/R/L/E/N across the 12 haplotypes; $P = 0.986$) and codon 475 (Ile566, four residues from the C-terminus; I/V/R/L/S/A/K; $P = 0.983$). Both are peripheral positions remote from the catalytic pocket (224 residues N-terminal and 66 residues C-terminal to the A-site, respectively); **neither lies within the inactivated A-site, and neither maps to an acidic surface patch** (see below).

The inactivated catalytic motif itself is the most strongly conserved part of the fossil protein, contradicting the prediction that its codons would be relaxed or positively selected. Partitioning the codon alignment and fitting M0 separately (same tree, F3x4 codon frequencies):

GGDEF partition (codon columns)	n codons	M0 ω	P (ω vs. 1)
Inactivated A-site 6-mer (DENLIG in reference)	6	0.0023	1.3×10^{-9}
Domain core (all other columns)	470	0.069	$< 10^{-127}$
Acidic surface-patch residues	20	0.845	0.42 ($\omega \approx 1$ not rejected)
Acidic patch excluding two A-site-adjacent columns	18	0.681	0.57

Across the 12 haplotypes the dead motif is conserved as D/E–D/E–N/D–L–I/F–G/A (position counts: D₅/E₇; D₉/E₃; N₁₀/D₂; L₁₂; I₁₀/F₂; G₁₀/A₂), with most within-position changes synonymous or biochemically conservative (acidic↔acidic, hydrophobic, Gly→Ala), and the fourth position (Leu) is invariant. Its ω (0.0023) is 30-fold below the domain core: after the ancestral, clade-specific inactivation substitutions were fixed (EDNLIG, DDDLIG, DENLIG, etc.), the resulting mutant motif has been under strong purifying selection ever since, rather than reverting or drifting toward neutrality. Conversely, the 20 acidic residues that constitute the DENLIG cytoplasmic surface patches (coordinates taken from the AF3-based electrostatics analysis, §3.1, and transferred through the alignment) evolve at $\omega \approx 0.85$, statistically indistinguishable from neutrality ($P = 0.42$; $\omega = 0.68$, $P = 0.57$ after excluding the two columns adjacent to the A-site)—roughly 12-fold higher than the core. This is the site-specific signature predicted for non-functional surface drift, and it is incompatible with purifying selection maintaining a metal-coordinating surface.

Interpretation. (i) Universal HD-GYP retention carries the expected molecular-evolutionary signature of adaptive maintenance—housekeeping-grade purifying selection ($M0 \omega = 0.008$)—directly supporting the Black Queen “retain the scavenging step” model (§3.2). (ii) The GGDEF fossil is not a neutrally decaying pseudogene: its TIM-barrel core ($\omega = 0.069$) and especially its inactivated motif ($\omega = 0.0023$) are under strong purifying selection, indicating the full-length membrane scaffold—and the specific mutant motif—retain a constrained structural role after catalytic loss. (iii) Relaxation of constraint is real but modest ($\omega \sim 4\text{--}8\times$ above recA/gyrB/rpoB and retained HD-GYP) and is spatially restricted: near-neutral evolution at acidic surface-patch residues ($\omega \approx 0.85$) and positive/relaxed diversification at two peripheral positions—not at the catalytic A-site. The pre-registered prediction of relaxed or positively selected catalytic glycine codons (§3.6) is therefore **not supported by the data**; the observed pattern instead matches “structural retention with peripheral surface drift.” (iv) Power caveats: genomes within each genus share near-identical proteins, leaving 12 and 6 unique haplotypes, and cross-genus synonymous-site saturation removes most pairwise comparisons (the gene-level M0 likelihood estimates use all taxa and are robust, but the pairwise Mann–Whitney tests are not); RELAX, branch-site, and FEL analyses (§4.6) have now been completed and are reported in the following paragraph.

HyPhy corroboration: no episodic positive selection, HD-GYP shows borderline intensification, and the two M8 BEB sites are not robust. Four HyPhy tests (BUSTED, RELAX, FEL, aBSREL; Methods §4.6) were run on the same six ortholog sets used for PAML. **BUSTED** (gene-wide episodic positive selection) finds no significant signal in any set ($P = 0.50$ for the GGDEF fossil and HD-GYP; $P = 0.051$ for recA—borderline but not significant at $\alpha = 0.05$; $P = 0.42$ for rpoB; $P = 0.50$ for dnaG; $P = 0.39$ for gyrB), confirming that no gene in the analysis, including the inactivated GGDEF fossil, has experienced gene-wide episodic positive selection. **RELAX** (relaxation parameter K, with all internal branches as the test set) gives $K = 1.09$ for the GGDEF fossil ($P = 0.374$, no relaxation or intensification), $K = 2.80$ for HD-GYP ($P = 0.111$, borderline intensification—directionally consistent with adaptive retention of HD-GYP but not reaching significance at the 0.05 level), and $K \approx 1$ for all four housekeeping controls (recA $K = 0.96$, rpoB $K = 1.00$, dnaG $K = 0.83$, gyrB $K = 1.12$; all $P > 0.05$). The HD-GYP $K > 1$ trend aligns with the PAML finding that HD-GYP is the most tightly constrained set ($M0 \omega = 0.0082$), but the borderline P-value means RELAX alone cannot formally distinguish intensification from neutrality at the 0.05 level. **FEL** (per-site dN/dS) on the GGDEF fossil directly tests the two M8 BEB-positive codons: site 180 ($\alpha = 1.41$, $\beta = 3.23$, $\omega = 2.30$, LRT = 0.24, **$P = 0.625$**) and site 475 ($\alpha \approx 0$, $\beta = 50.8$, $\omega \rightarrow \infty$, LRT = 2.55, **$P = 0.110$**)—neither reaches significance at $P < 0.05$. The M8 positive-selection signal at these two peripheral positions is therefore not robust to a per-site likelihood-ratio test, and the GGDEF fossil shows no site-level positive selection that survives the test. Across all 476 codons of the GGDEF fossil, FEL identifies 218 sites under significant purifying selection ($P < 0.05$; all with $\omega \ll 1$), consistent with the M0 global estimate ($\omega = 0.067$). **aBSREL** (per-branch episodic selection) on the GGDEF fossil finds zero branches with significant episodic positive selection, again consistent with BUSTED and FEL. Together, the four HyPhy tests corroborate the PAML interpretation: the GGDEF fossil is under global purifying selection with no robust positive-selection signal at any site or branch; HD-GYP shows a borderline intensification trend ($K > 1$) consistent with adaptive retention; and the two M8 BEB sites should be treated as false positives of the site model rather than as evidence for localized positive selection.

2.11 Cross-phylum signal-domain inventory: HMMER-verified comparison

To place Kryptonion’s c-di-GMP pathway degradation in a broader evolutionary context, we scanned the complete proteomes of four reference organisms—*Aquifex aeolicus* VF5 (Aquificae), *Thermus thermophilus* HB8 (Deinococcus–Thermus), *Escherichia coli* K12 (Proteobacteria), and *Pyrococcus* sp. (Euryarchaeota)—against the same 12-Pfam signal-domain HMM database used for Kryptonion (Methods §4.2), using the Pfam gathering threshold. The Kryptonion row reports unique proteins after within-phylum deduplication; genome-occurrence figures (HD-GYP 51/51, PilZ 3/51) are from §2.2–2.4.

Organism (proteome size)	Phylum	GGDEF (PF00990)	EAL (PF00563)	HD (PF01966)	HisKA (PF00512)	Response_reg (PF00072)	PilZ (PF07238)	c-di-GMP system status
<i>Kryptonion</i> (51 genomes Kryptonion pooled)		1†	0	13	26	45	0	Synthesis lost; HD-GYP retained (51/51 genomes)
<i>Aquifex aeolicus</i>	Aquificae	0	0	7	8	13	0	Complete pathway absence

Organism (proteome size)	Phylum	GGDEF (PF00990)	EAL (PF00563)	HD (PF01966)	HisKA (PF00512)	Response_reg (PF00072)	PilZ (PF07238)	c-di- GMP system status
VF5 (1,626 proteins) <i>Thermus thermophilus</i> HB8 (2,258 proteins)	Deinococcus– Thermus	3	1	12	10	14	0	Reduced but intact pathway
<i>E. coli</i> K12 (4,376 proteins)	Proteobacteria	18	17	6	21	37	2	Full c-di- GMP system
<i>Pyrococcus</i> sp. (~1,800 proteins)	Euryarchaeota	0	0	5	1	0	0	No c-di- GMP system

†The single Kryptonia GGDEF HMM hit is the lone active GGDEF singleton (MCX7761429.1 in GCA_026418375.1); all 44 inactive GGDEF domains have accumulated sufficient catalytic-site substitutions (EDNLIG, DDDLIG, DENLIG, EDNLFA, SAHSFG) that they no longer meet the PF00990 gathering threshold, providing an independent HMM-based confirmation of catalytic-site degradation beyond the motif-based classification.

The HMMER-verified comparison confirms four points. (i) *Aquifex aeolicus* lacks both GGDEF and EAL entirely (0/0), matching Kryptonia’s synthesis-loss pattern—but unlike Kryptonia, Aquifex has also lost HD-GYP at the pathway level (the 7 HD-domain proteins detected do not carry the canonical HD-GYP[DE] C-terminal signature in Kryptonia’s pattern), so the complete c-di-GMP absence in Aquifex is attributable to extreme genome reduction (1,626 proteins), a different mechanism from Kryptonia’s asymmetric “synthesis-lost / degradation-retained” pattern. (ii) *Thermus thermophilus* retains a reduced but intact c-di-GMP system (3 GGDEF, 1 EAL, 12 HD), demonstrating that thermophily alone does not drive c-di-GMP loss—consistent with the broader literature survey of seven thermophile phyla (§3.4). (iii) *E. coli* K12 retains the full c-di-GMP signaling module (18 GGDEF, 17 EAL, 2 PilZ), representing the canonical complete system. (iv) *Pyrococcus* (archaeon) lacks the entire c-di-GMP system, as expected for Archaea. These HMMER-verified counts provide direct computational support for the cross-phylum claims in §3.4 and confirm that Kryptonia’s asymmetric degradation pattern is not a generic thermophile trait.

3. Discussion

3.1 The catalytically inactive GGDEF molecular fossil

The near-universal conservation of catalytically inactive GGDEF motifs across **51 Kryptonia genomes**—encoded by full-length 555–589 aa membrane proteins—is the most striking finding of this study. Six motif variants were identified (EDNLIG, DDDLIG, EDNLFA, DENLIG, SAHSFG, and a single GGDEF), with 44 of 45 GGDEF domains (97.8%) carrying non-Gly substitutions at BOTH catalytic glycine positions. Several lines of evidence indicate these are **molecular fossils** rather than sequencing artifacts or recent pseudogenes: (i) the proteins are full-length and not truncated; (ii) the inactivating substitutions at both catalytic glycine positions are conserved across all species and all five genera, indicating loss prior to phylum radiation; (iii) motif variation is clade-specific (EDNLIG/EDNLFA in North American genomes, DDDLIG in JGI genomes, DENLIG in Malaysian genomes), indicating that the inactivating substitutions were fixed early, at or before the phylum radiation; codon-based analysis (§2.10) further shows that within each clade the resulting mutant motif has subsequently been maintained under strong purifying selection ($\omega = 0.0023$), so the motif is not undergoing continuous neutral decay; (iv) SRA read-mapping at 6–19× coverage independently confirmed the catalytic codons in representative genomes (GGDEF active, DENLIG and DDDLIG inactive), ruling out assembly artifacts; (v) the proteins retain membrane-integral architecture, suggesting they may have been exapted for a non-catalytic function (e.g., protein–protein interaction or structural scaffolding); (vi) **AlphaFold3 structure predictions of six representative GGDEF domains confirm that all inactive variants retain the canonical (β/α)₈ TIM-barrel fold** ($C\alpha$ -RMSD 5.0–6.1 Å vs. active GGDEF) with conserved catalytic-site backbone geometry ($C\alpha$ – $C\alpha$ distance 3.81–3.91 Å), demonstrating that the fold is stably maintained despite catalytic-site ablation. Critically, the overall fold is universally retained, and codon-based analysis (§2.10) confirms that both the fold core ($\omega = 0.069$) and—unexpectedly—the inactivated A-site motif itself ($\omega = 0.0023$, the most strongly constrained region of the entire protein) are maintained under purifying selection; the allosteric I-site (RxxD) is completely degenerated in all inactive domains (0/44; only the single active GGDEF retains a canonical I-site), and statistically detectable relaxation is spatially restricted to acidic surface-patch residues of the DENLIG/EDNLIG clade ($\omega \approx 0.85$, indistinguishable from neutrality; the clade shows a modest $\sim 1.3\times$ acidity increase over the proteome background). We interpret the surface shift as surface-property drift accompanying functional loss, not as evidence for metal-binding or exaptation for scavenging. These motifs therefore represent catalytic-site fossils whose structural scaffold is stably retained (a purifying-selection signature now directly demonstrated, §2.10), joining a small but growing list of enzymes that mark ancient pathway loss while their non-catalytic scaffold persists.

Structural basis of the molecular fossil. The structural predictions provide mechanistic insight into how catalytic function is lost while the fold is preserved. In the active GGDEF domain, the two catalytic glycines (Gly1 and Gly2 of the GG[D/E]EF A-site) form part of the GTP-binding pocket; their small side chains create the steric space required for the guanine base and ribose to dock. In all five inactive variants, both glycines are replaced by bulkier residues (Glu/Asp/Ser at position 1; Asp/Glu/Ala at position 2), which sterically occlude the nucleotide-binding site. Because these substitutions occur at surface-exposed loop positions rather than within the hydrophobic core, they do not destabilize the (β/α)₈ barrel—explaining why the fold remains intact (pLDDT 65–75) while catalysis is abolished. This is the structural signature of a molecular fossil: a precise, chemically targeted inactivation of the active site with preservation of the global fold. The fact that the inactivating substitutions are confined to the two catalytic glycines—rather than being scattered throughout the domain—suggests that the loss of catalysis was driven by selection (e.g., energetic cost of c-di-GMP synthesis in oligotrophic conditions) while the fold itself was preserved for an alternative function, consistent with exaptation rather than neutral pseudogenization. The dN/dS results (§2.10) strengthen this reading in two ways: the inactivated motif codons are the most strongly conserved in the entire protein ($\omega = 0.0023$), which is incompatible with ongoing neutral decay of the catalytic

loop and implies that the mutant motif itself fulfills a structural constraint; and the two codons with a positive-selection signature ($\omega \approx 4.6$) lie at distant peripheral positions (residues 271 and the C-terminal tail at 566), not at the catalytic site.

From molecular fossil to neutral scaffold: surface drift, not metal-binding exaptation. Two structural observations were evaluated for exaptation, but neither supports a scavenging function. First, the allosteric **I-site (RxxD motif)**—which in active DGCs binds the c-di-GMP product for feedback inhibition—is **completely degenerated in all 44 inactive domains (0/44)**; only the single active GGEEF retains a canonical I-site (RxxD at −9 aa upstream of the A-site). RxxD motifs found at non-canonical positions in DENLIG (−62 aa) and SAHSFG (−55/−38 aa) represent random background, not a conserved allosteric site. The I-site therefore provides **no evidence** for c-di-GMP binding by any inactive GGDEF domain. Second, the surface of inactive DENLIG/EDNLIG domains is moderately more acidic than the active GGEEF (D+E ~16–18% vs. ~15% of all residues in the catalytic domain; only ~1.3× over the proteome background of 12.4%), with up to five acidic patches and a seven-residue acidic spatial cluster (seven carboxylate side chains within 12 Å; not a contiguous sequence run). This acidic shift is **clade-specific**: DENLIG and EDNLIG show enhanced acidity; SAHSFG, DDDLIG, and EDNLFA show reduced or neutral acidity. We do **not** interpret the acidic surface as metal-binding evidence, for five reasons: (i) the enrichment is weak (~1.3× over background); (ii) none of the six GGDEF proteins carry a metal-domain HMM hit (the PFAM metalloproteome re-scan of 3,021 proteins contains zero GGDEF proteins), so there is no independent support for a metal-binding role; (iii) the excess acidic residues concentrate near the substituted catalytic pentapeptide, consistent with drift at the inactivated active site rather than selection for a metal-binding surface; (iv) **membrane topology analysis localizes all acidic patches to the cytoplasmic face of the membrane**. Transmembrane helix prediction using DeepTMHMM and TMHMM (DTU Health Tech), cross-validated with local Kyte–Doolittle hydrophobicity scanning and the positive-inside rule, shows that the C-terminal GGDEF domains of all three representative proteins reside on the cytoplasmic side. Specifically, in DENLIG (576 aa) the final transmembrane helix ends at residue 55 and all 23 acidic-patch residues (positions 419–570) are cytoplasmic; in EDNLIG (574 aa) the final transmembrane helix ends at residue 49 and all 19 acidic-patch residues (positions 417–568) are cytoplasmic; in GGEEF (320 aa) the sole transmembrane helix ends at residue 23 and all 8 acidic-patch residues (positions 200–317) are cytoplasmic. No acidic-patch residue maps to the extracellular/periplasmic side in any protein. Because cytoplasmic free metal-ion concentrations are tightly buffered at sub-micromolar levels and mineral precipitation does not occur in the cytoplasm, a metal-scaffold or extracellular mineralization function is not supported; (v) **three-dimensional coordination geometry analysis confirms that the acidic patches do not form a metal-binding site**. We compared the carboxylate oxygen geometry in the DENLIG GGDEF domain against two genuine Kryptonia metal-binding sites: (a) the heme-Fe coordination sphere in a multi-heme cytochrome c (AF3 model), where each heme Fe is coordinated by 4–6 ligands at 1.5–2.1 Å (Fe–N_{porphyrin} 1.98–2.05 Å, Fe–N_{His} 1.5–1.9 Å, Fe–S_{Met} 2.06 Å), with ligand–Fe–ligand angles of 88–92° (cis) and 177–180° (trans), matching octahedral/square-pyramidal geometry; and (b) the HD-GYP phosphodiesterase active site (AF3 models of 2 Kryptonia HD-GYP proteins, MCS7229639 and MGC8595393, both carrying the canonical HD-GYP[DE] catalytic signature: HD motif followed by GYPE at 25-residue spacing with multiple coordinating Glu residues), where the catalytic HD motif positions the His imidazole ND1 nitrogen and the adjacent Asp carboxylate oxygen (OD1/OD2) at 5.2–7.6 Å apart—ideal for sandwiching one or two metal ions (2.0–2.3 Å from each ligand) in the bimetallic Fe/Mn center. The two HD-GYP active sites show rich mixed-ligand composition: MCS7229639 contributes 4 His imidazole N, 10 Asp carboxylate O, and 14 Glu carboxylate O ligand atoms (total 28); MGC8595393 contributes 4 His, 6 Asp, and 16 Glu ligand atoms (total 26), with additional downstream Asp/Glu residues (e.g., Glu40, Glu42, Glu43, Glu57, Glu74, Glu75, Glu80 in MCS7229639; Glu45, Glu47, Glu49, Glu62, Glu68, Glu79, Glu80, Glu85 in MGC8595393) completing the coordination sphere and providing the mixed His/Asp/Glu ligand set characteristic of HD-GYP metalloenzymes. In DENLIG, the catalytic site (residue 494) has only 4 carboxylate O atoms within 10 Å, all from 1–2 Asp/Glu residues with no His ligand; the tightest carboxylate–O cluster (residues 419, 421) has O atoms 1.76–3.18 Å from the cluster centroid—too dispersed to coordinate a metal ion (genuine sites require all ligand atoms within 2.0–2.5 Å of the metal, plus a mixed His/Asp/Glu ligand set rather than carboxylates alone). Pocket geometry analysis further rules out a solvent-accessibility explanation: the DENLIG catalytic site has a pocket openness of only 0.025 (fraction of 200 random directions from the centroid reaching solvent within 6 Å)—tighter than HD-GYP (mean openness 0.153, range 0.035–0.270). The atom packing density around the DENLIG catalytic site (0.0499 atoms/Å³ within 8 Å) is comparable to HD-GYP (0.0317–0.0504 atoms/Å³), and the convex-hull pocket volume (373.1 Å³) falls within the HD-GYP range (185.3–432.4 Å³), indicating a well-packed, sealed pocket with a volume comparable to genuine metal-binding sites rather than a shallow groove. The metal-binding deficiency is therefore a **ligand-composition and geometry defect**—absence of His ligands and dispersed carboxylate oxygens—rather than a pocket-shape or solvent-exposure defect. None of the three carboxylate–O clusters in DENLIG satisfies the metal-coordination criterion, despite the seven-residue acidic spatial cluster (within 12 Å). The acidic surface therefore lacks both the geometric prerequisite and the ligand-type diversity required for metal binding. The inactivated GGDEF domain is therefore best described as a **neutral structural fossil**: the catalytic machinery is discarded, but the membrane-anchored scaffold persists (demonstrated to be under purifying selection, §2.10), with clade-specific surface drift that carries no functional implication. Codon analysis has now tested the corresponding prediction directly (§2.10): the 20 acidic-patch residues evolve at $\omega \approx 0.85$, statistically indistinguishable from neutrality ($P = 0.42$; $\omega = 0.68$, $P = 0.57$ after excluding the two columns adjacent to the A-site)—12-fold more relaxed than the purged domain core ($\omega = 0.069$)—whereas the inactivated A-site motif itself is, unexpectedly, under even stronger purifying selection than the core ($\omega = 0.0023$). Neutral drift of the acidic surface is thus confirmed while a metal-binding exaptation requiring conserved ligand geometry is not.

The single active GGEEF: a secondary acquisition. The one exception to phylum-wide inactivation—*Candidatus* Kryptonium sp. GCA_026418375.1, which encodes both an active GGEEF and an inactive DENLIG—actually strengthens the molecular-fossil interpretation. The DENLIG protein in this genome is the canonical inactivated fossil shared with all other *Kryptonium* DENLIG genomes (GCA_025060635.1, GCA_033580455.1, GCA_039630395.1). The additional GGEEF protein, which has no counterpart in any other Kryptonia genome, most plausibly represents a **secondary acquisition by horizontal gene transfer** that occurred after the ancestral inactivation event. Under this scenario, the phylum-wide GGDEF inactivation is a derived, fixed trait, and the single active copy is a lineage-specific novelty rather than a retained ancestral feature. This interpretation is testable: phylogenetic analysis of the GGEEF protein should place it outside the Kryptonia GGDEF clade, and compositional analysis (GC content, codon usage) may reveal a foreign signature.

3.2 HD-GYP retention: a molecular footprint of Black Queen dynamics

The **asymmetric retention of HD-GYP phosphodiesterases (PF01966 superfamily in 51/51 genomes, 100%; canonical HD-GYP[DE] in 46/51, 90%)** in the absence of active diguanylate cyclase synthesis is a particularly telling finding. EAL-type PDEs are completely absent (0/51), yet HD-GYP PDEs—also capable of c-di-GMP degradation—are retained across all five genera. This pattern cannot be explained by neutral drift alone: if c-di-GMP signaling were entirely lost, both PDE families would be expected to decay. The near-universal retention of HD-GYP suggests a **functional role in degrading environmental c-di-GMP** leaked by community helpers (e.g., *Fervidibacter sacchari*). Kryptonia thus acts as a “cheater” that exploits c-di-GMP as a public good: it has lost the costly synthesis machinery but retains the degradation apparatus, potentially using environmental c-di-GMP as a signal of community density or helper presence. This provides direct genomic evidence for the Black Queen Hypothesis operating on intracellular signaling molecules.

Evolutionary significance of HD-GYP retention. Three considerations deepen the interpretation of this finding:

1. *Why HD-GYP and not EAL?* EAL and HD-GYP are both c-di-GMP phosphodiesterases but differ in catalytic mechanism and substrate range. EAL domains typically require an EAL→EAL dimeric active site and hydrolyze c-di-GMP to pGpG with high specificity. HD-GYP domains, in contrast, are metal-dependent ($\text{Fe}^{2+}/\text{Mn}^{2+}$) hydrolases that can process both c-di-GMP and its linear derivative pGpG, and some HD-GYP enzymes exhibit broader phosphodiesterase activity toward additional cyclic dinucleotides. In the oligotrophic, metal-rich hot-spring environment, the **metal-dependent HD-GYP mechanism may be energetically cheaper** than the dimeric EAL mechanism, and its broader substrate range may allow Kryptonia to scavenge multiple nucleotide second messengers leaked by the community. This would explain why HD-GYP, not EAL, was retained during streamlining.

Structural comparison of HD-GYP and EAL phosphodiesterases. The two PDE families diverge fundamentally in fold, active-site architecture, and catalytic chemistry, which directly bears on why HD-GYP was retained and EAL lost in Kryptonia:

Feature	EAL domain (PF00563)	HD-GYP domain (PF01966)
Fold	$(\beta/\alpha)_8$ TIM-barrel (modified)	α/β sandwich (HD superfamily)
Active-site residues	EAL triad → EAL motif (Glu-Ala-Leu) + conserved Asp/Glu ligands; catalytic Glu activated by $\text{Mg}^{2+}/\text{Mn}^{2+}$	H...D...GYP motif; His-Asp dyad coordinates 1–3 metal ions ($\text{Fe}^{2+}/\text{Mn}^{2+}$); terminal Tyr assists leaving group
Metal requirement	Requires $\text{Mg}^{2+}/\text{Mn}^{2+}$; obligate dimeric active site (two EAL domains face each other)	Requires $\text{Fe}^{2+}/\text{Mn}^{2+}$; can be monomeric or dimeric; active site forms within a single domain
Dimerization	Obligate homodimer for activity (EAL–EAL interface builds the catalytic pocket)	Facultative ; many HD-GYPs are active as monomers
Substrate range	Narrow: c-di-GMP → pGpG (linear dinucleotide); typically does not hydrolyze pGpG further	Broader: c-di-GMP → pGpG → 2 GMP; can also act on c-di-AMP and other cyclic dinucleotides
Energy cost of activity	Continuous Mg^{2+} occupancy + dimerization interface maintenance	Metal ions scavenged from environment; no obligate dimer
Thermal stability	Dimer interface is often the thermolabile element; EAL PDEs in thermophiles require compensatory stabilizing mutations	Compact α/β fold with metal-core stabilization; intrinsically more thermostable

Three structural rationales account for HD-GYP's selective retention: - **(i) Monomeric competence.** Unlike EAL, which must dimerize to form a catalytic pocket, HD-GYP can function as a monomer. In a streamlined genome with reduced protein-expression capacity, a single-domain monomeric enzyme imposes lower biosynthetic and folding burden than an obligate homodimer. - **(ii) Metal-rich habitat matching.** Hot springs are rich in Fe^{2+} and Mn^{2+} , the preferred cofactors of HD-GYP. EAL relies on Mg^{2+} , which is also abundant, but HD-GYP's ability to use Fe^{2+} directly couples catalysis to the geochemical milieu without additional metal-import cost. - **(iii) Degradation–recovery chain to GMP/GTP.** HD-GYP opens c-di-GMP to the linear dinucleotide pGpG; a separate calcineurin-like metallo-phosphatase (PF00149, 302 hits) then hydrolyzes pGpG to GMP, which guanylate kinase (gmk) and nucleoside-diphosphate kinase (ndk) convert to GDP and GTP. This complete, enzyme-validated chain (§2.8) delivers usable nucleotide nutrition from the leaked public good. (We note that HD-GYP may in some organisms terminate at pGpG, which itself could serve as a phosphate source; the further processing to GMP/GTP is catalyzed by the dedicated phosphodiesterase and kinases.)

2. *Evolutionary sequence of pathway disintegration.* The genomic fossil record suggests a **stepwise degradation** of the c-di-GMP pathway: (i) the ancestral Kryptonium possessed a complete DGC+EAL/HD-GYP+PilZ system; (ii) the first inactivating substitution occurred in the GGDEF catalytic motif, abolishing synthesis while the protein retained full length and membrane localization; (iii) under relaxed selection, PilZ receptors decayed first (retained in only 6% of genomes), followed by EAL PDEs (0% retention); (iv) **HD-GYP PDEs were universally retained**, which we hypothesize reflects adaptive retention for degrading environmental c-di-GMP—rather than being degraded. Codon-based analysis has now confirmed the predicted selection asymmetry (§2.10): the retained HD-GYP evolves under housekeeping-grade purifying selection ($M0 \omega = 0.0082$; pairwise median 0.010), 8.2-fold tighter than the inactivated GGDEF fold ($\omega = 0.067$) and statistically indistinguishable from gyrB ($P = 0.83$); the pre-specified one-sided HD-GYP < GGDEF pairwise test falls narrowly above the 0.05 threshold ($P = 0.065$) because only two informative HD-GYP pairs survive dS-saturation filtering among the six unique haplotypes.

3. *Cheater fitness benefit.* For an obligate-fermentative, phosphorus-limited thermophile, retaining HD-GYP confers two fitness advantages at negligible cost: (a) it opens environmental c-di-GMP to pGpG, which the downstream degradation–recovery chain (metallo-phosphatase → GMP → gmk → GDP → ndk → GTP, §2.8) converts into usable nucleotide nutrition—turning a leaked public good into a phosphorus and nucleotide resource; (b) it enables **density sensing**: by degrading c-di-GMP at the cell surface, Kryptonia may titrate the local concentration of this molecule and infer community helper density, gating growth or attachment decisions accordingly. The near-total loss of PilZ receptors (3/51) is consistent with this model: Kryptonia no longer needs to *respond* to its own c-di-GMP (which it cannot make), but it benefits from *clearing* environmental c-di-GMP for its degradation products and as a sensory input.

Together, these considerations position HD-GYP retention as a **paradigm case of adaptive gene retention under the Black Queen Hypothesis**: a component of a lost pathway is not merely a decaying relic but is actively maintained because it has been exapted to exploit the very public good whose production the cell has abandoned.

Nucleotide nutrition recovery: the third axis of the cheater strategy. The retention of HD-GYP must be understood not in isolation but as one end of a **complete nucleotide-recovery pipeline** whose other components—the downstream degradation enzymes and the uptake machinery—are retained across the phylum. The chain has three layers. **(i) Extracellular cleavage.** HD-GYP (37 canonical HD-GYP[DE] proteins across 46/51 genomes) opens leaked c-di-GMP to pGpG at the cell surface, where a calcineurin-like metallo-phosphatase (PF00149, 323 hits across 51 genomes) hydrolyzes it to GMP. This is the only step unique to the c-di-GMP public good and the only one tied to the lost signaling pathway. **(ii) Uptake.** Kryptonia encodes an unusually large nucleoside-transport complement—CNT/NupC (39), Nucleoside_tra2 (146), and ABC substrate-binding proteins (767) under the original broader-family scan, totaling 952 transporter proteins (Nucleoside_tra2 = 69 across 45 genomes and ABC = 170 across 51 genomes in the conservative 8-profile re-scan; §2.8)—far exceeding what a streamlined fermenter would need for its own salvage alone. This mirrors the oligopeptide-uptake strategy described for amino acids (§2.9): Kryptonia does not maintain biosynthetic pathways for leaky metabolites but instead imports them as pre-formed units (oligopeptides for amino acids; nucleosides/nucleotides for purines). **(iii) Intracellular channeling.** Guanylate kinase (gmk, 147 PF02779 family hits across 51 genomes) and nucleoside-diphosphate kinase (ndk, 47 hits

across 47 genomes) phosphorylate the imported GMP through GDP to GTP, feeding the guanine nucleotide pool directly. The net effect is that **a discarded signaling molecule becomes a phosphorus and purine resource**: each molecule of extracellular c-di-GMP delivers two guanine bases and four phosphate groups—comparable in phosphorus content to two ATP molecules—recovered at the cost of one HD-GYP and two kinases, with no expenditure on de novo purine synthesis. This places nucleotide recovery firmly within the same BQH logic as amino-acid and cofactor auxotrophy (§2.9): leaky, community-suppliable resources are scavenged rather than synthesized. Three caveats apply. First, this is a functional inference from pathway completeness; direct experimental demonstration is required. Second, Kryptonita is not wholly dependent on scavenged nucleotides: de novo purine synthesis (e.g., *guaB*/IMP dehydrogenase, PF00478 = 100 hits across 50 genomes in the re-scan; §2.8) is retained, so recovery supplements rather than replaces synthesis. Third, HD-GYP need not drive the chain all the way to GTP; if degradation terminates at pGpG, the linear dinucleotide may itself be imported (Nucleoside_tra2) or serve as a phosphate source without full phosphorylation. Regardless of the terminal step, the presence of all three layers—cleavage, uptake, and channeling—establishes nucleotide recovery as a coherent, genome-encoded trait rather than a coincidental by-product of enzyme promiscuity.

3.3 An energy-economics model for graded signal degradation

We propose that the graded loss of nucleotide second-messenger pathways reflects an **energy-economics trade-off** in the oligotrophic hot-spring environment:

Pathway	Energetic cost	Essentiality	Fate
			Catalysis lost; GGDEF scaffold retained as neutral structural fossil (A-site inactivated in 44/45 domains; I-site degenerated in all inactive domains [0/44]; fold preserved; modest acidic surface drift in DENLIG/EDNLIG, not interpreted as metal-binding)
c-di-GMP synthesis	2 GTP/mol + continuous turnover	Low (biofilm, motility; community-suppliable)	
c-di-GMP degradation (HD-GYP)	No ATP/GTP; metal-dependent hydrolysis	Medium–High (environmental c-di-GMP scavenging → GMP phosphorus recovery; density sensing)	Retained (PF01966 100%; HD-GYP[DE] 90%)
cAMP	1 ATP/mol	Medium (carbon metabolism; CRP retained)	Semi-retained
ppGpp	1 ATP+GTP/mol (transient burst)	High (stringent response to starvation)	Retained
TCS	Phosphoryl transfer (no nucleotide cofactor)	High (environmental sensing)	Retained

c-di-GMP synthesis is the most energetically expensive second-messenger reaction: each molecule requires two GTP, and robust signaling demands continuous synthesis and degradation. In a phosphorus- and energy-limited hot spring, this cost is prohibitive. Because c-di-GMP can diffuse across membranes and be shared as a community public good, individual cells can lose the synthesis pathway and exploit community-supplied c-di-GMP (BQH). The retention of HD-GYP PDEs—which require no nucleotide cofactor for catalysis—allows Kryptonita to degrade environmental c-di-GMP into GMP at minimal energetic cost, recovering phosphorus and providing a sensory readout of helper density. cAMP, requiring only one ATP and needed for carbon-source regulation, is semi-retained with a functional receptor. ppGpp and TCS are essential for nutrient-stress response and environmental sensing and cannot be lost.

3.4 Implications for the Black Queen Hypothesis and the phylum-level degradation framework

By analyzing 51 genomes—49 from hot springs in North America and Asia plus two non-hot-spring European MAGs (§4.1)—and spanning all five candidate genera, this study establishes that the observed signal-pathway degradation is not an idiosyncrasy of one isolate or one community but a **phylum-wide degradation pattern**. The pattern is reproducible across all five candidate genera (*Kryptonium*, *Kryptoniota*, *Kryptobacter*, *Thermokryptus*, *Chryseopegocella*) and across geographically separated populations, indicating that the degradation trajectory was set in the common ancestor of the phylum and has been maintained by selection ever since.

Cross-phylum comparison situates the pattern. A survey of c-di-GMP pathway inventories across seven thermophile phyla (HMMER-verified comparison in §2.11; literature survey in Table S1) places Kryptonita’s pattern in a broader evolutionary context. The intact c-di-GMP systems of *Thermus thermophilus* (Deinococcus–Thermus), *Thermotoga maritima* (Thermotogae), *Dictyoglomus thermophilum* (Dictyoglomi), and *Geobacillus stearothermophilus* (Firmicutes) demonstrate that **thermophily alone is insufficient to drive c-di-GMP loss**. The complete pathway absence in *Aquifex aeolicus* (Aquificae) is attributable to extreme genome reduction (1,626 proteins), a different mechanism. Only *Caldiserica* (*Caldisericum exile*) shares Kryptonita’s asymmetric “synthesize-lost / HD-GYP-retained” signature—both are obligately fermentative, genome-streamlined thermophiles from oligotrophic hot springs. This convergent pattern across two unrelated phyla suggests that the **graded signal degradation pattern is driven by the combined selective pressure of obligate fermentation, genome streamlining, and oligotrophy**, rather than by phylogeny or thermophily. Kryptonita is distinguished by the completeness and universality of the degradation (44/45 GGDEF domains inactive across all 51 genomes and 5 genera, with a single active GGEEF in one genome; 100% HD-GYP retention), with no lineage showing reversion to active synthesis from the ancestral fossil—consistent with a single ancestral inactivation event fixed by selection and maintained by BQH dynamics.

Within this global framework, the study extends BQH in five ways:

1. **Signaling as a leaky function.** BQH has primarily been demonstrated for metabolic public goods (vitamins, amino acids). We show that **intracellular signaling molecules** (c-di-GMP) can also function as leaky public goods, with pathway loss compensated by community supply. Critically, c-di-GMP is not actively secreted—both Kryptonia and the Aquificae control lack all dedicated secretion systems (T1SS–T6SS). Candidate extracellular routes are passive leakage from lysed cells and vesicle-associated transfer: c-di-GMP is known to regulate OMV biogenesis (*Pseudoalteromonas marina*; Hu et al., 2024; Wu et al., 2023), while direct c-di-GMP cargo measurements in community OMVs remain to be done. Either route yields a naturally diffusible public good exploitable by cheaters.
2. **Graded loss, not binary.** Rather than a simple retain/lose dichotomy, we observe a **degradation gradient** (c-di-GMP synthesis < c-di-GMP degradation/HD-GYP < cAMP < ppGpp/TCS) that correlates with cost and essentiality. This suggests BQH operates along a continuum rather than as an all-or-nothing process.
3. **Structural vs. signaling metalloprotein dichotomy.** Within the same metalloproteome, structural metalloproteins (hemes, Fe-S clusters, zinc ribbons) are retained while signaling metalloproteins (GGDEF, EAL) are lost. This demonstrates that BQH operates at the sub-proteome level: cell-autonomous structural functions are protected, while community-shareable signaling functions are dispensable.
4. **Adaptive retention of degradation in a lost pathway.** The universal retention of HD-GYP PDEs (100% of genomes) after loss of synthesis demonstrates that BQH can produce **asymmetric pathway remnants**: the cell jettisons the costly, shareable step (synthesis) while retaining the low-cost, resource-recovering step (degradation). This refines BQH from a simple “lose the leaky function” model to a more nuanced “lose the costly step, retain the scavenging step” model. The single active GGEEF in GCA_026418375.1—likely a secondary HGT acquisition—does not contradict this model; rather, it demonstrates that the phylum-wide inactivation is sufficiently fixed that even a newly acquired active DGC coexists with the inactivated fossil without reversing the degradation trajectory.
5. **Neutral retention of the inactivated scaffold, not exaptation.** BQH has traditionally framed gene loss as a passive process: the cell discards a costly, shareable function and relies on the community. Our structural analysis shows that the inactivated GGDEF scaffold is **retained but not exapted for scavenging**: the allosteric I-site (RxxD) is completely degenerated in all inactive domains (0/44), so c-di-GMP binding via the I-site is not supported; the modest acidic surface shift in DENLIG/EDNLIG (~1.3× over background) is not interpreted as metal-binding evidence (no metal-domain HMM hits, enrichment too weak). The codon analysis (§2.10) adds the decisive observation: the scaffold persists under strong purifying selection on the fold core and the mutant motif ($\omega = 0.069$ and 0.0023), while near-neutral evolution is confined to the acidic surface patches ($\omega \approx 0.85$, indistinguishable from 1)—i.e., selection preserves the architecture but does not preserve any catalytic or metal-binding geometry. This refines the BQH model from “loss-and-exaptation” to “loss-and-neutral-structural-retention”: the degraded enzyme scaffold is maintained as a constrained structural element without acquiring a new resource-scavenging role.

3.5 The zinc-ribbon connection: signal coupling retained at the structural level

The Cys3-His1 configuration of the DnaG ZBD is notable. While Kryptonia abandons c-di-GMP as a regulatory signal, it retains a structural zinc site that directly couples environmental Zn^{2+} concentration and pH to DNA replication. This represents a **regulatory strategy shift**: from diffusible second-messenger signaling (lost) to direct metal-ion/redox sensing at the replication apparatus (retained). The His residue in the ZBD coordination sphere introduces pH sensitivity, potentially allowing the cell to gate replication in response to hot-spring pH fluctuations.

3.6 Limitations and future directions

This study is based on comparative genomic and metagenomic inference, with structural evidence derived from **AlphaFold3 computational predictions** (not experimentally determined structures). The structural conclusions—fold retention, catalytic-site backbone geometry, I-site positioning, and surface electrostatics—should be interpreted as computational corroboration of the sequence-based motif analysis rather than as experimentally determined structures; high-resolution structural validation remains a future goal. Several predictions require experimental validation: (i) the inactivated GGDEF protein is catalytically inactive; (ii) c-di-GMP is undetectable in Kryptonia cells; (iii) cAMP pathway genes are transcriptionally responsive to cAMP; (iv) gene_12976 assembles [4Fe-4S] clusters and mediates electron transfer; (v) Fervidibacter supplies c-di-GMP or amino acids to Kryptonia; (vi) the original prediction—HD-GYP under purifying selection with a relaxed inactivated A-site—**was tested (§2.10)**: retained HD-GYP is under housekeeping-grade purifying selection ($\omega = 0.0082$) while the fossil GGDEF is modestly relaxed overall ($\omega = 0.067$), but the inactivated A-site itself is the *most* constrained region ($\omega = 0.0023$)—the specific prediction of relaxed/positive catalytic codons is not supported; the predicted neutrality of the DENLIG/EDNLIG acidic surface patches is supported ($\omega \approx 0.85$, $P = 0.42$ vs. $\omega = 1$); (vii) HD-GYP can hydrolyze environmental c-di-GMP into GMP that feeds the guanine salvage pathway; (viii) the inactivated DENLIG/EDNLIG GGDEF domain does **not** bind Fe^{2+}/Mn^{2+} at the acidic surface patches; the evolutionary half of this prediction is now satisfied—the patches carry no signature of purifying selection ($dN/dS \approx 1$) even though the domain core and mutant motif are strongly conserved—and c-di-GMP binding via the I-site remains unsupported (the I-site is degenerated in all inactive domains); (ix) community-derived OMVs directly carry c-di-GMP as cargo: current evidence only establishes that c-di-GMP regulates OMV biogenesis in *Pseudoalteromonas marina* (Hu et al., 2024; Wu et al., 2023), whereas cargo transfer in this community is the untested step of the public-good model. Experimental validation is planned and will be reported separately.

dN/dS analysis: outcome and remaining analyses. The codon-based test has been executed (Results §2.10; Methods §4.6). Outcome relative to the pre-registered predictions: (a) **supported**—retained HD-GYP is under housekeeping-grade purifying selection ($M0 \omega = 0.0082$; pairwise median 0.010 , comparable to *gyrB*); (b) **partially supported**—the inactivated GGDEF is modestly relaxed overall ($\omega = 0.067$; significantly above *recA*, *gyrB*, *rpoB* but not *dnaG*) yet remains under strong global purifying selection, so it is not a neutrally decaying pseudogene; (c) **not supported**—the prediction that the two catalytic A-site positions are relaxed or positively selected is rejected: the inactivated 6-mer is the most constrained part of the protein ($\omega = 0.0023$), and the only M8 BEB-positive codons (positions 271 and 566) are peripheral; (d) **supported**—the DENLIG acidic surface-patch residues evolve near neutrally ($\omega = 0.85/0.68$, indistinguishable from 1), whereas the fold core is purged ($\omega = 0.069$), consistent with surface drift and inconsistent with metal-binding exaptation. The pre-specified one-sided Mann–Whitney comparison (HD-GYP < GGDEF) gave $P = 0.065$ rather than $P < 0.05$: effective power is far below the originally assumed 44-vs-51-sequence design because within-genus proteins are $\geq 99\%$ identical (12 and 6 unique haplotypes after dereplication) and cross-genus synonymous sites are saturated (2 and 14 informative pairwise comparisons). **HyPhy corroboration.** RELAX, FEL, and BUSTED/aBSREL (HyPhy) have now been completed on the same six ortholog sets and corroborate the PAML interpretation (Results §2.10): (i) BUSTED finds no gene-wide episodic positive selection in any set (all $P \geq 0.05$; *recA* borderline at $P = 0.051$); (ii) RELAX gives $K = 1.09$ for the GGDEF fossil ($P = 0.374$, no relaxation) and $K = 2.80$ for HD-GYP ($P = 0.111$, borderline intensification consistent with adaptive retention but not reaching significance); (iii) FEL directly tests the two

M8 BEB-positive codons in the GGDEF fossil—site 180 ($P = 0.625$) and site 475 ($P = 0.110$)—neither reaches significance, so the M8 positive-selection signal at these peripheral positions is not robust to a per-site likelihood test. aBSREL on the GGDEF fossil finds zero branches with episodic positive selection. PAML branch-site model A was attempted but could not be completed due to tree-file parsing issues in codeml with the labeled Newick format; however, BUSTED (gene-wide episodic selection) and aBSREL (per-branch episodic selection) together answer the equivalent question—no episodic positive selection is detectable in the GGDEF fossil. The remaining limitation is that all analyses use FastTree LG+ γ guide trees rather than species-level IQ-TREE/ModelFinder trees; this does not affect M0 or partition estimates but could influence per-site FEL localization.

4. Materials and Methods

Scope note. Sections 4.1–4.6 describe analyses **executed in this study** at an overview level; every numerical claim in Results is reproducible from the consolidated data files listed in Data Availability. Detailed experimental parameters, tool versions, and analysis scripts will be supplied in the journal-submission version. Section 4.7 refers to planned experimental validation: no protocol details, methods, or data from it appear in this manuscript, and results will be reported separately upon completion.

4.1 Genomes and annotation

51 Kryptonion genomes were obtained from NCBI Assembly (including GCA_061246185.1, a complete genome), selected for assembly quality ($N50 \geq 10$ kb) and spanning all five candidate genera from global hot-spring locations (habitat audit below). These represent the subset of 120 high-quality Kryptonion assemblies for which protein annotations were available in NCBI. Proteins were predicted and annotated using standard microbial-genome tools (ab initio gene calling, Pfam HMMER domain scan, KEGG KO assignment, and InterProScan). Control genomes (*Thermus thermophilus* HB8, *Aquifex aeolicus* VF5, *E. coli* K12) were analyzed identically.

Habitat composition and assembly-quality audit. BioSample metadata confirm that 49/51 genomes derive from hot springs (Tengchong and other Chinese sites, $n = 29$; Dewar Creek, Canada, $n = 12$; Malaysia, $n = 4$; Russia, $n = 2$; USA, $n = 1$; Japan, $n = 1$). The two exceptions are non-hot-spring MAGs: GCA_036567965.1 (wetland soil metagenome, Czech Republic) and GCA_049278975.1 (freshwater sediment metagenome, Denmark). Both are statistical outliers in genome metrics—protein counts 3,327 and 3,034 versus 1,331–2,514 in the other 49 genomes; assembly sizes 4.15 and 3.36 Mb versus 1.57–2.88 Mb; GC 50.0% and 45.5% versus 36.5–42.5%—and GCA_049278975.1 is additionally a highly fragmented assembly (329 contigs, $N50 = 14.3$ kb). Neither genome carries an NCBI contamination or RefSeq-exclusion flag, and both pass phylum-marker checks. We retained both genomes because every core conclusion is demonstrably insensitive to them: excluding both leaves GGDEF inactivation at 42/43 domains (97.7%), HD-GYP retention at 49/49 (100%), I-site degeneration at 0/42, and the SAHSFG motif census at 2 genomes. The genome-wide census statistics that do include them are reported for transparency, and readers should treat per-genome totals for these two accessions with caution.

4.2 Signal domain inventory

Pfam HMMs used: PF00990 (GGDEF), PF00563 (EAL), PF01966 (HD), PF07238 (PilZ), PF00211 (Guanylate_cyc / Class III AC), PF01928 (CYTH / Class IV AC), PF00512/PF07568/PF07730 (HisKA), PF00072 (REC), PF00027 (cNMP-binding), PF00672 (HAMP).

GGDEF catalytic motif extraction (corrected method). The catalytic A-site pentapeptide was identified using the HMMER domain envelope rather than the protein C-terminus, because GGDEF domains may be followed by additional domains. Within each GGDEF domain, the catalytic site was pinpointed from the conserved profile–target alignment. This avoids the pitfall of a loose whole-domain regex search, which can pick up non-catalytic glycines (e.g., I-site residues, structural glycines). A motif was classified as active only if it matched GG[D/E]EF; all other sequences were classified inactive. This method identified 45 GGDEF domains across 44 genomes; 44 (97.8%) carry non-Gly substitutions at both catalytic glycine positions, while a single domain in GCA_026418375.1 retains the canonical active GGEEF motif.

SRA read-mapping validation of catalytic motifs. For representative genomes with available SRA runs, raw sequencing reads were mapped to the corresponding genome assemblies. For each GGDEF protein, the gene coordinates were determined, the coding sequence extracted and translated to locate the catalytic A-site motif, and per-base pileup was performed at the catalytic codon positions. A motif was considered confirmed if the majority base call at each catalytic codon position matched the reference assembly, supported by ≥ 3 reads. Three motif variants were validated: GGEEF (active, GCA_026418375.1, 6 $\times$ coverage), DENLIG (inactive, GCA_026418375.1, 3–4 $\times$), and DDDLIG (inactive, GCF_001458535.1, 18–19 $\times$).

Cross-phylum HMMER-verified comparison. Complete proteomes of four reference organisms—*Aquifex aeolicus* VF5, *Thermus thermophilus* HB8, *Escherichia coli* K12, and *Pyrococcus* sp.—were scanned against the same 12-Pfam signal-domain HMM database used for Kryptonion. Unique target proteins per domain were counted. The Kryptonion pooled-proteome scan was performed identically; within-phylum near-identical proteins are counted once per unique sequence.

4.3 AlphaFold3 structure prediction of GGDEF domains

Target selection. Six GGDEF-domain proteins were selected for structure prediction to span the catalytic-motif diversity and taxonomic breadth: one active GGEEF (GCA_026418375.1, MCX7761429) and five inactive variants—DENLIG (GCA_026418375.1, MCX7762467; GCA_025060635.1, MCS7229517), EDNLIG (GCF_001485605.1, *Kryptobacter*), DDDLIG (GCF_001536205.1, *K. thompsonii*), and SAHSFG (GCA_036567965.1, *Kryptoniota*).

Multiple sequence alignment (MSA) construction. For each target, a custom MSA was built using Jackhmmer against a combined database comprising the SwissProt FASTA database and a curated set of Kryptonion GGDEF-domain sequences.

Structure prediction. AlphaFold3 was run with the generated MSA and default model parameters. For each target, the best-ranked model (by model confidence) was selected for analysis. Model confidence was assessed using the predicted local distance difference test (pLDDT) score per residue and the predicted aligned error (PAE) matrix.

Structural analysis. The GGDEF catalytic domain was defined as residues flanking the catalytic-motif position (covering the $(\beta/\alpha)_8$ TIM-barrel core). α -backbone superposition was performed with the active GGEF structure as the reference, and root-mean-square deviation (RMSD) was computed over the aligned catalytic-domain α atoms. The α – α distance between the first two catalytic-motif residues was measured to assess backbone geometry conservation at the active site.

I-site (RxxD allosteric site) identification. The canonical GGDEF I-site (RxxD motif) was identified by regular-expression scanning of each protein sequence, and the distance (in residues) between each RxxD match and the catalytic A-site was recorded. An RxxD motif was classified as a **canonical I-site** only if it fell within the **–15 to –5 residue window upstream of the A-site** (corresponding to the structurally characterized I-site position, ~9 residues upstream, where the Arg side chain contacts the c-di-GMP phosphate backbone). RxxD motifs outside this window were recorded as non-canonical background occurrences.

Surface electrostatics and hydrophobicity analysis. Surface-exposed residues were identified from α coordination, and surface acidity was quantified as the fraction of surface residues that are aspartate (D) or glutamate (E). Acidic patches were defined as spatial clusters of acidic surface residues; the size of the largest such cluster was recorded. Surface hydrophobicity was computed as the mean hydropathy score over surface-exposed residues. All analyses were computed within the GGDEF catalytic domain.

Membrane topology analysis. To determine the subcellular side on which the GGDEF catalytic domains (and their acidic patches) reside, transmembrane topology was predicted for representative full-length proteins using independent deep-learning and hidden-Markov-model predictors, cross-validated with hydrophobicity scanning and the positive-inside rule. The subcellular localization of each acidic-patch residue was determined by mapping its sequence position to the predicted topology map.

4.4 Structural metalloprotein inventory and coordination geometry analysis

Sequence-based metal-binding motif scan. All 106,651 proteins across the 51 genomes were scanned for 14 known metal-binding sequence motifs using regular expressions: EF-hand (DxDxDG, Ca^{2+}), C2H2/C4/CCHC/C3H1 zinc fingers (Zn^{2+}), zinc ribbon (CxxC...CxxC, Zn^{2+}), heme c (CXXCH, Fe), HExxH (Zn^{2+}), ferritin diiron (E...E...E...H, Fe), Walker A (G...GK[TS], Mg^{2+}), rubredoxin (CxxC...CxxC, Fe), cupin (G-x_s-H-x-H, Fe/Zn/Mn), 4Fe-4S (CxxCxxC...C, Fe-S), and 2Fe-2S (C-x₂-C...C...C, Fe-S). For Fe-S clusters, custom spacing rules were applied (4th Cys within 30 aa for 4Fe-4S; 15–30 aa and 1–5 aa spacing for 2Fe-2S). Per-motif protein counts, total hit counts, genome coverage, and ligand-spacing statistics were computed. Hits were cross-validated against HMMER Pfam domain assignments (PF00990, PF00563, PF01966, etc.) to confirm protein-family membership.

PFAM HMMER structural metalloprotein inventory. To independently quantify structural metalloprotein families by domain homology (rather than short sequence motifs), all 106,651 proteins across the 51 genomes were scanned with a panel of PFAM hidden Markov models covering eight structural metalloprotein categories (EF-hand, c-type cytochromes, zinc ribbons, Rieske 2Fe-2S, [4Fe-4S] clusters, heme non-c, copper proteins, and Fe-S clusters). A protein was assigned to a category if it harbored at least one significant hit from any HMM in that category; the total is the non-redundant count of proteins with at least one metalloprotein-category hit.

Three-dimensional coordination geometry analysis. To test whether the GGDEF acidic patches could form a metal-binding site, we compared the carboxylate-oxygen geometry in the DENLIG GGDEF AlphaFold3 model against genuine Kryptonite metal-binding sites: the heme-Fe coordination sphere in a multi-heme cytochrome c model and the HD-GYP phosphodiesterase active site. For each structure, side-chain metal-ligand atoms (His, Asp, Glu, heme N) were extracted and pairwise distances, centroid-ligand distances, and ligand–metal–ligand angles were computed.

Pocket geometry analysis. To distinguish a ligand-composition defect from a solvent-exposure defect, we computed pocket-geometry metrics around the catalytic-site centroid (atom packing density, pocket volume, and pocket openness) and compared them between HD-GYP and the DENLIG GGDEF domain.

Read-back validation of metagenomic multi-heme cytochromes. Raw paired-end reads for the local hot-spring community sample were mapped to the non-redundant local assembly, and per-base depth was computed. For each target cytochrome the coding sequence was located on its contig and mean depth, zero-coverage fraction, and per-motif depth at all CXXCH positions were recorded. For gene_8068 the CDS had a mean depth of 19.5 $\times$ (20.3 $\times$ against the single contig as sole reference), 0.0% zero-coverage positions, and all 17 CXXCH motifs covered at $\geq 5\times$; gene_14270 and gene_3950 CDSs had mean depths of 31.8 $\times$ and 84.0 $\times$, respectively. CXXCH motifs were detected with the regular expression $\text{C}[\text{ }^{\sim}\text{C}]{2}\text{CH}$ (X_2 excluding Cys, the canonical heme c-binding spacing) unless stated otherwise.

4.5 Community helper–cheater analysis

The hot-spring metagenome was binned and taxonomically classified. GGDEF/EAL/HD-GYP counts were tallied per taxon. Helper species were defined as carrying both a functional GGDEF and a functional EAL/HD-GYP. Cheater species were defined as lacking functional PDE despite retaining (degraded) GGDEF.

4.6 Molecular-evolution (dN/dS) analysis

Codon-based selection analyses were performed on the 51-genome dataset to test whether HD-GYP retention is adaptive and to localize any relaxation of constraint within the inactivated GGDEF protein.

- **Ortholog set construction.** Candidate proteins were taken from the GGDEF and HD-domain HMM hits supplemented by motif regular expressions, and four housekeeping controls (recA, gyrB, rpoB, dnaG) were retrieved by annotation. All-vs-all BLASTp was followed by symmetric bitscore-ratio reciprocal clustering with identity and coverage gates; connected components were defined by a disjoint-set union. The resulting families span 37 genomes (inactivated GGDEF; the lone active GGEF protein is a one-member component and was excluded), 42 genomes (HD-GYP core family), and 36–40 genomes (each housekeeping control).
- **Dereplication.** Each family was dereplicated at high amino-acid identity to prevent pseudoreplication of identical sequences and zero-length branches, leaving 12 GGDEF, 6 HD-GYP, and 8–9 housekeeping unique haplotypes for model fitting; all genome memberships are retained in the manifest.

- **Alignment and back-translation.** Unique haplotypes were aligned with MAFFT (L-INS-i algorithm), and codon alignments were generated with pal2nal in gap-stripping mode using the corresponding genomic CDS sequences.
- **Phylogeny.** Guide trees were inferred with FastTree (LG + Γ approximation) from the amino-acid alignments. Zero-length branches and multifurcations were handled by flooring internal branch lengths and binarizing polytomies. These trees serve as fixed topologies for codon models; a species-level phylogenetic analysis is listed among future analyses (§3.6).
- **Selection inference.** For every family: (a) a one-ratio model (M0) with ω free and with ω fixed at 1, giving the per-gene ω and a likelihood-ratio test of purifying selection; (b) models M7 (beta) and M8 (beta + one free ω class), compared by likelihood-ratio test, with Bayes Empirical Bayes sites recorded; (c) pairwise estimates of per-pair dN, dS, and ω . For the GGDEF family, partitioned M0 runs compared the inactivated A-site codons, the acidic-patch residues, and the remaining core codons.
- **Statistics.** Pairwise ω values with saturated dS were excluded before computing medians/IQRs; one-sided Mann–Whitney U tests compared the focal sets against housekeeping controls. Likelihood-ratio test p-values used the appropriate χ^2 distributions.
- **HyPhy corroboration.** Four HyPhy tests were run on the same ortholog sets: BUSTED (gene-wide episodic positive selection), RELAX (relaxation/intensification parameter K), FEL (per-site dN/dS), and aBSREL (per-branch episodic selection on the GGDEF fossil). These corroborate the PAML interpretation. PAML branch-site model A was attempted but could not be completed due to tree-file parsing issues; BUSTED and aBSREL together answer the equivalent question.

4.7 Planned functional validation

Experimental validation of the predictions listed in §3.6 is planned; the corresponding methods and results will be reported separately upon completion. No protocol details, methods, or data from these experiments appear in this version.

5. Conclusions

We present a comparative genomic and metagenomic panorama of the Kryptonia phylum across **51 genomes (49 from hot springs in North America and Asia, plus two non-hot-spring European MAGs; §4.1), the annotated subset of 120 high-quality assemblies and the largest annotated Kryptonia dataset to date**, establishing a **phylum-wide degradation pattern** of nucleotide second-messenger degradation: c-di-GMP synthesis is effectively lost (inactivated GGDEF molecular fossil in 44/45 domains across all five genera; a single active GGDEF in one genome likely represents secondary HGT acquisition) while HD-GYP degradation is universally retained (100% of genomes) to scavenge environmental c-di-GMP; cAMP is semi-retained (Class IV CyaB + functional CRP, 100% conserved); and ppGpp plus two-component systems are fully intact (100% conserved). SRA read-mapping at 6–19 $\times$ coverage independently confirmed representative catalytic motifs (GGDEF active, DENLIG and DDDLIG inactive). **AlphaFold3 structure predictions of six GGDEF domains (one active + five inactive across three genera) provide direct structural corroboration of the molecular-fossil hypothesis:** all inactive domains retain the canonical (β/α)₈ TIM-barrel fold ($C\alpha$ -RMSD 5.0–6.1 Å vs. active GGDEF) with conserved catalytic-site backbone geometry; the allosteric I-site (RxxD) is completely degenerated in all inactive domains (0/44; only the active GGDEF retains a canonical I-site) while the catalytic A-site is universally destroyed; and inactive DENLIG/EDNLIG surfaces show a modest clade-specific acidification (D+E ~16–18% vs. ~15%, ~1.3 $\times$ over proteome background), which we report as surface-property drift accompanying functional loss and do not interpret as metal-binding evidence (weak enrichment, zero metal-domain HMM hits, cytoplasmic localization of all acidic patches, and absence of metal-binding coordination geometry confirmed by 3D comparison with genuine heme-Fe and HD-GYP metal sites; pocket geometry analysis confirms a sealed catalytic pocket (openness 0.025 vs. HD-GYP mean 0.153; convex-hull volume 373.1 vs. HD-GYP 185.3–432.4 Å³), ruling out a solvent-exposure or pocket-shape explanation). Structural metalloproteins—[4Fe-4S] clusters, zinc ribbons, and multi-heme cytochromes—are stably retained. **Codon-based selection analysis across the 51 genomes (§2.10) adds the molecular-evolutionary dimension to this model:** the universally retained HD-GYP evolves under housekeeping-grade purifying selection ($\omega = 0.008$), directly supporting its adaptive maintenance; the inactivated GGDEF TIM-barrel ($\omega = 0.069$) and, unexpectedly, its mutant catalytic motif itself ($\omega = 0.0023$, the most constrained region of the protein) are also under strong purifying selection, whereas near-neutral evolution is confined to the acidic surface-patch residues ($\omega \approx 0.85$, indistinguishable from 1) and two peripheral positions flagged by PAML M8 BEB but not robustly supported by per-site FEL ($P = 0.625$ and 0.110)—refuting the predicted relaxation at the catalytic codons and refining the fossil concept to “structurally constrained scaffold with neutral surface drift.” HyPhy BUSTED, RELAX, FEL, and aBSREL (§2.10) corroborate this interpretation: no gene-wide, site-level, or branch-level episodic positive selection is detectable in the GGDEF fossil, and HD-GYP shows a borderline intensification trend ($K = 2.80$, $P = 0.111$) consistent with adaptive retention. Community analysis identifies *Fervidibacter sacchari* as the sole c-di-GMP helper, with Kryptonia as a cheater. We propose that this pattern reflects an energy-economics trade-off in oligotrophic hot springs, extending the Black Queen Hypothesis to the level of intracellular signaling and demonstrating that costly, shareable pathway steps are selectively lost while low-cost, resource-recovering steps are adaptively retained.

Supplementary Tables

Table S1. The 51 Kryptonia genomes analyzed in this study. All accessions are from NCBI Assembly; these 51 genomes are the annotated subset (protein FASTA available) of 120 high-quality Kryptonia assemblies (N50 ≥ 10 kb). Species assignments follow NCBI taxonomy. Habitats (NCBI BioSample metadata, §4.1): 49/51 are from hot springs; the two exceptions are the non-hot-spring MAGs GCA_036567965.1 (#24, wetland soil, Czech Republic) and GCA_049278975.1 (#25, freshwater sediment, Denmark), which are also outliers in protein count, assembly size, and GC content (§4.1 audit).

#	Accession	Species	N50 (bp)	Proteins
1	GCA_061246185.1	<i>Candidatus</i> Kryptonium sp.	2,528,525	2,275
2	GCA_025060635.1	<i>Candidatus</i> Kryptonium sp.	174,405	2,113
3	GCA_026418375.1	<i>Candidatus</i> Kryptonium sp.	65,147	2,069
4	GCA_033580455.1	<i>Candidatus</i> Kryptonium sp.	171,320	1,969

#	Accession	Species	N50 (bp)	Proteins
5	GCA_039630395.1	<i>Candidatus</i> Kryptonium sp.	25,779	2,376
6	GCF_001458535.1	<i>Candidatus</i> Kryptonium thompsonii	26,952	2,045
7	GCF_001458575.1	<i>Candidatus</i> Kryptonium thompsonii	25,049	1,859
8	GCF_001485625.1	<i>Candidatus</i> Kryptonium thompsonii	342,558	2,274
9	GCF_001535965.1	<i>Candidatus</i> Kryptonium thompsonii	57,692	2,116
10	GCF_001535985.1	<i>Candidatus</i> Kryptonium thompsonii	34,662	1,989
11	GCF_001536005.1	<i>Candidatus</i> Kryptonium thompsonii	23,715	2,096
12	GCF_001536025.1	<i>Candidatus</i> Kryptonium thompsonii	24,564	1,823
13	GCF_001536085.1	<i>Candidatus</i> Kryptonium thompsonii	24,228	1,993
14	GCF_001536125.1	<i>Candidatus</i> Kryptonium thompsonii	18,419	1,710
15	GCF_001536145.1	<i>Candidatus</i> Kryptonium thompsonii	50,594	2,231
16	GCF_001536185.1	<i>Candidatus</i> Kryptonium thompsonii	18,283	1,331
17	GCF_001536205.1	<i>Candidatus</i> Kryptonium thompsonii	18,789	1,836
18	GCF_001458495.1	<i>Candidatus</i> Kryptobacter tengchongensis	17,628	1,716
19	GCF_001458555.1	<i>Candidatus</i> Kryptobacter tengchongensis	17,711	1,566
20	GCF_001485605.1	<i>Candidatus</i> Kryptobacter tengchongensis	241,734	2,325
21	GCF_001536105.1	<i>Candidatus</i> Kryptobacter tengchongensis	12,528	1,568
22	GCF_001536065.1	<i>Candidatus</i> Chryseopegocella kryptomonas	102,670	1,901
23	GCF_900070205.1	<i>Candidatus</i> Thermokryptus mobilis	138,435	2,235
24	GCA_036567965.1	<i>Candidatus</i> Kryptoniota bacterium	178,152	3,327
25	GCA_049278975.1	<i>Candidatus</i> Kryptoniota bacterium	14,325	3,034
26	GCA_049966155.1	<i>Candidatus</i> Kryptoniota bacterium	256,836	2,309
27	GCA_049967335.1	<i>Candidatus</i> Kryptoniota bacterium	44,426	2,514
28	GCA_055840995.1	<i>Candidatus</i> Kryptoniota bacterium	49,048	2,092
29	GCA_055856755.1	<i>Candidatus</i> Kryptoniota bacterium	297,646	2,274
30	GCA_055891375.1	<i>Candidatus</i> Kryptoniota bacterium	659,885	2,323
31	GCA_055896195.1	<i>Candidatus</i> Kryptoniota bacterium	10,574	1,889
32	GCA_055896355.1	<i>Candidatus</i> Kryptoniota bacterium	34,048	2,201
33	GCA_055896815.1	<i>Candidatus</i> Kryptoniota bacterium	13,991	1,838
34	GCA_055896855.1	<i>Candidatus</i> Kryptoniota bacterium	29,376	2,034
35	GCA_055903035.1	<i>Candidatus</i> Kryptoniota bacterium	11,726	1,952
36	GCA_055908875.1	<i>Candidatus</i> Kryptoniota bacterium	42,920	2,176
37	GCA_055909035.1	<i>Candidatus</i> Kryptoniota bacterium	118,887	2,317
38	GCA_055928005.1	<i>Candidatus</i> Kryptoniota bacterium	119,368	2,224
39	GCA_055929125.1	<i>Candidatus</i> Kryptoniota bacterium	241,692	1,799
40	GCA_055929405.1	<i>Candidatus</i> Kryptoniota bacterium	44,839	2,139
41	GCA_055953565.1	<i>Candidatus</i> Kryptoniota bacterium	28,976	1,937
42	GCA_056001685.1	<i>Candidatus</i> Kryptoniota bacterium	19,463	1,885
43	GCA_056004445.1	<i>Candidatus</i> Kryptoniota bacterium	452,576	2,333
44	GCA_056007485.1	<i>Candidatus</i> Kryptoniota bacterium	63,604	2,128
45	GCA_056007825.1	<i>Candidatus</i> Kryptoniota bacterium	167,380	2,266
46	GCA_056025325.1	<i>Candidatus</i> Kryptoniota bacterium	10,011	2,132
47	GCA_056027485.1	<i>Candidatus</i> Kryptoniota bacterium	10,158	1,862
48	GCA_056064075.1	<i>Candidatus</i> Kryptoniota bacterium	19,198	2,083
49	GCA_056076055.1	<i>Candidatus</i> Kryptoniota bacterium	20,265	2,037
50	GCA_056077955.1	<i>Candidatus</i> Kryptoniota bacterium	12,987	2,085
51	GCA_056078455.1	<i>Candidatus</i> Kryptoniota bacterium	15,995	2,045

Table S2. Catalytic-motif variants of the 45 Kryptonia GGDEF domains. The catalytic A-site was identified using the HMMER domain envelope and the conserved ggde match state within the GGDEF (PF00990) HMM alignment. Both catalytic glycine positions are non-Gly in 44/45 domains.

Motif	n	Position 1	Position 2	Status	Genus distribution
EDNLIG	15	Glu (E)	Asp (D)	inactive	<i>Kryptoniota</i> (N. America)
DDDLIG	11	Asp (D)	Asp (D)	inactive	<i>Kryptonium thompsonii</i> (JGI)
EDNLFA	9	Glu (E)	Asp (D)	inactive	<i>Kryptoniota</i> (N. America)
DENLIG	5	Asp (D)	Glu (E)	inactive	<i>Kryptonium</i> (Malaysia)
SAHSFG	4	Ser (S)	Ala (A)	inactive	<i>Kryptoniota</i>
GGEEF	1	Gly (G)	Gly (G)	active	<i>Kryptonium</i> sp. GCA_026418375.1

Table S3. Signal-domain inventory across 51 Kryptonia genomes. Counts are from Pfam HMMER (E-value < 0.01, domE < 0.01).

Domain	Pfam ID	Genomes with ≥ 1 hit	Total copies	Mean copies/genome
GGDEF	PF00990	44/51 (86%)	45	0.88
GGDEF_2	PF13675	0/51 (0%)	0	0
EAL	PF00563	0/51 (0%)	0	0
HD (HD-GYP)	PF01966	51/51 (100%); canonical HD-GYP[DE] 46/51 (90%)	455 (37 canonical HD-GYP[DE])	8.92
PilZ	PF07238	3/51 (6%)	3	0.06
HisKA	PF00512	51/51 (100%)	1,208	23.7
Response_reg	PF00072	51/51 (100%)	1,588	31.1
cNMP_binding	PF00027	51/51 (100%)	290	5.69
HAMP	PF00672	51/51 (100%)	322	6.31
CYTH (CyaB)	PF01928	51/51 (100%)	102	2.00
Guanylate_cyc (Class III AC)	PF00211	51/51 (100%)	204	4.00

Table S4. Multi-heme c-type cytochromes with ≥ 8 CXXCH motifs in the local hot-spring metagenome (GY_2023_03 / MbPL202605402). CXXCH motifs detected with the regular expression $C[\text{ }^{\sim}C]\{2\}CH$ (X_2 excluding Cys). Read-back coverage is per-base depth (BWA-MEM, 51.8 M raw read pairs, samtools depth -a); the “TPM-based” values are the community-sample transcript-abundance estimate. All eight ORFs are complete.

Gene	CXXCH Length (aa)	Contig: CDS	Strand	Mean read-back depth	
gene_8068	17	654	k127_6393:409–2373	+	19.5× (≈12× TPM-based)
gene_16296	13	386	k127_4885:285–1445	–	15.5×
gene_14270	11	423	k127_17347:1–1269	–	31.8× (10.2× TPM-based)
gene_6340	9	562	k127_3560:510–2198	–	27.8×
gene_22171	9	268	k127_10500:3–806	+	26.0×
gene_1608	9	406	k127_9374:359–1579	+	22.0×
gene_3950	8	653	k127_2903:1161–3122	–	84.0× (6.5× TPM-based)
gene_26034	8	475	k127_16941:262–1689	–	29.6×

Note: gene_3950 is the 653-aa locally assembled core domain of the 823-aa NCBI entry WP_143714178.1 (annotated “partial”); the two share identical eight-CXXCH spacing.

Supplementary Figures

Figure S2. Structural superposition of six GGDEF domains: molecular-fossil evidence. (A) α -backbone cartoon superposition of the six predicted GGDEF catalytic domains (residues ± 80 of the catalytic motif), with the active GGEEF (MCX7761429, orange) as the reference and the five inactive variants shown in distinct colors (DENLIG $\times 2$ in blue/cyan, EDNLIG in green, DDDLIG in magenta, SAHSFG in yellow). All inactive domains retain the canonical (β/α)₈ TIM-barrel fold despite non-Gly substitutions at both catalytic positions. [Panel A rendered from AlphaFold3 mmCIF coordinate files; see Data Availability.] (B) Distribution of α – α distances between the first two catalytic-motif residues across the six structures. All distances fall within 3.81–3.91 Å (active and inactive alike), demonstrating conserved catalytic-site backbone geometry. (C) Catalytic glycine substitution map: in the active GGEEF, Gly1 and Gly2 (orange spheres) create the steric space for GTP docking; in each inactive variant, both glycines are replaced by bulkier residues (Glu/Asp/Ser at position 1; Asp/Glu/Ala at position 2) that occlude the nucleotide-binding pocket without disrupting the barrel fold. α -RMSD values relative to GGEEF are 5.0–6.1 Å over 155–161 aligned residues. All structures were predicted by AlphaFold3 with Jackhmmer-built MSAs (mean pLDDT = 65–75). Quantitative metrics in panels B–C (α – α distances, pLDDT, motif residues) are verified from ggdef_structure_comparison_6.json.

Figure S3. I-site degeneration and surface electrostatics of inactivated GGDEF domains: molecular-fossil drift signatures. (A) Canonical I-site (RxxD at ~ 9 aa upstream of the A-site) retention: only the active GGEEF domain (1/6 structures, 1/45 genome-wide) retains a canonical I-site; all 44 inactive domains lack a canonical I-site. RxxD motifs found in DENLIG (–62 aa) and SAHSFG (–55/–38 aa) are at non-canonical positions. (B) Surface electrostatic potential map (red = acidic, blue = basic) of the active GGEEF and representative inactive DENLIG/EDNLIG domains, showing enhanced negative surface charge in the inactive clades. [Rendered from AlphaFold3 mmCIF coordinate files; see Data Availability.] (C) Acidic (D/E) and basic (R/K) surface residue fraction by motif: DENLIG (24.3%/22.6%) and EDNLIG (21.5%) exceed the active GGEEF (19.3%), whereas SAHSFG (15.9%) is reduced. Note: these are surface-residue fractions; the all-residue D+E fraction is ~ 16 –18% for DENLIG/EDNLIG vs. ~ 15 % for GGEEF ($\sim 1.3\times$ over proteome background). (D) Net surface charge (acidic % – basic %) by motif: DENLIG (+5.2 to +6.4) and EDNLIG (+2.9) are strongly electronegative relative to GGEEF (+0.1), while SAHSFG (–2.4) is electropositive. (E) Surface hydrophobicity (Kyte–Doolittle) comparison across motifs. (F) Drift signature matrix: canonical I-site count (RxxD at –5 to –15 aa), number of acidic patches (≥ 3 acidic residues within 12 Å), and largest acidic-cluster size by motif. DENLIG has ≥ 5 acidic patches but no canonical I-site; EDNLIG has acidic patches only; SAHSFG/DDDLIG/EDNLFA have neither canonical I-site nor enhanced acidity. The acidic patches are not interpreted as metal-binding evidence (no metal-domain HMM hits; enrichment $\sim 1.3\times$ over background; all acidic patches localize to the cytoplasmic face by membrane topology analysis; 3D coordination geometry comparison with genuine heme-Fe and HD-GYP metal sites confirms the carboxylate-O clusters are too dispersed and lack the mixed His/Asp/Glu ligand set required for metal coordination; pocket geometry analysis confirms a sealed catalytic pocket (openness 0.025 vs. HD-GYP mean 0.153; convex-hull volume 373.1 vs. HD-GYP 185.3–432.4 Å³), ruling out a solvent-exposure or pocket-shape explanation). (G) Schematic of the “factory smashed, gripper lost” model: active GGEEF (green) has both A-site (GTP synthesis) and I-site (allosteric regulation); ancestral inactivation ablates the A-site ($\times$) and degenerates the I-site ($\times$) in all inactive domains; the DENLIG/EDNLIG clade shows a modest acidic surface drift (red dots) that is not interpreted as metal-ion binding (cytoplasmic localization + sealed pocket with no metal-binding ligand composition/geometry). All quantitative surface analyses (C–F)

References

1. Eloe-Fadrosh EA, et al. (2016). Global metagenomic survey reveals a new bacterial candidate phylum in geothermal springs. *Nat Commun* 7:10476.
2. Morris JJ, Lenski RE, Zinser ER (2012). The Black Queen Hypothesis: evolution of dependencies through adaptive gene loss. *mBio* 3(2):e00036-12.
3. Hengge R (2009). Principles of c-di-GMP signalling in bacteria. *Nat Rev Microbiol* 7:263–273.
4. Moser CC, Keske JM, Warncke K, Farid RS, Dutton PL (1992). Nature of biological electron transfer. *Nature* 355:796–802. (Source of the ≤ 14 Å edge-to-edge biological electron-tunneling ruler cited in §2.6.)
5. Figueroa-Gonzalez P, et al. (2020). Peptide uptake as a strategy for amino acid acquisition in *Saccharibacteria*. *Front Microbiol* 11:588362.
6. Nou NO, Covington JK, Lai D, Mayali X, Seymour CO, Johnston J, Jiao JY, Buessecker S, Mosier D, Muok AR, Torosian N, Cook AM, Briegel A, Woyke T, Eloe-Fadrosh E, Shapiro N, Bryan SG, et al., Hedlund BP (2024). Genome-guided isolation of the hyperthermophilic aerobic *Fervidibacter sacchari* reveals conserved polysaccharide metabolism in the *Armatimonadota*. *Nat Commun* 15:9534. doi:10.1038/s41467-024-53784-3. (Cited for the taxonomic identity, aerobic hyperthermophilic physiology, and polysaccharide-degrading ecology of *F. sacchari*; the 7-GGDEF/1-EAL “sole helper” census in §2.8 is this study’s own metagenomic count, not a claim of ref. 6.)
7. Jumper J, et al. (2021). Highly accurate protein structure prediction with AlphaFold. *Nature* 596:583–589.
8. Abramson J, et al. (2024). Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature* 630:493–500.
9. Eddy SR (2011). Accelerated profile HMM searches. *PLoS Comput Biol* 7(10):e1002195.
10. Mistry J, et al. (2021). Pfam: The protein families database in 2021. *Nucleic Acids Res* 49(D1):D412–D419.
11. Li H, Durbin R (2009). Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* 25(14):1754–1760.
12. Danecek P, et al. (2021). Twelve years of SAMtools and BCFtools. *GigaScience* 10(2):giab008.
13. Katoh K, Standley DM (2013). MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol* 30(4):772–780.
14. Minh BQ, et al. (2020). IQ-TREE 2: New models and efficient methods for phylogenetic inference in the genomic era. *Mol Biol Evol* 37(5):1530–1534.
15. Yang Z (2007). PAML 4: phylogenetic analysis by maximum likelihood. *Mol Biol Evol* 24(8):1586–1591.
16. Suyama M, Torrents D, Bork P (2006). PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments. *Nucleic Acids Res* 34:W609–W612.
17. Price MN, Dehal PS, Arkin AP (2010). FastTree 2—approximately maximum-likelihood trees for large alignments. *PLoS ONE* 5(3):e9490.
18. Camacho C, et al. (2009). BLAST+: architecture and applications. *BMC Bioinformatics* 10:421.
19. Hu X-M, Peng L, Wu J, Wu G, Liang X, Yang J-L (2024). Bacterial c-di-GMP signaling gene affects mussel larval metamorphosis through outer membrane vesicles and lipopolysaccharides. *npj Biofilms Microbiomes* 10. doi:10.1038/s41522-024-00508-6. (Experimental evidence in *Pseudoalteromonas marina* that deleting the GGDEF/EAL gene *cdgB* raises cellular c-di-GMP ($\sim 17.5\%$) and increases OMV ($\sim 86\%$) and EPS (~ 3 -fold) production; c-di-GMP–OMV correlation $r = 0.73$, $P = 0.002$. Establishes regulation of OMV biogenesis by c-di-GMP; does not assay c-di-GMP as OMV cargo.)
20. Wu G, Hu X-M, Peng L-H, Zhang W, Liang X, Dobretsov S, Yang J-L (2023). Outer membrane vesicles induce the mussel plantigrade settlement via regulation of c-di-GMP. *Biofouling* 39(4):359–370. doi:10.1080/08927014.2023.2215707.

Data Availability Statement

The 51 Kryptonia genomes analyzed here are available from NCBI Assembly (accessions in Table S1). The local hot-spring metagenome raw reads and the non-redundant assembly are being prepared for deposition in a public archive (NCBI SRA under a new BioProject); accessions will be provided in the journal-submission version. In the interim, the raw reads and all derived data tables are available from the corresponding author upon reasonable academic request. AlphaFold3 structural outputs are archived at Zenodo (DOI 10.5281/zenodo.23015087). Custom analysis scripts and the derived data tables described below are available from the corresponding author upon reasonable academic request and will be released in a public repository at journal submission. The inactivated GGDEF catalytic-motif alignment and signal-domain count matrix are provided in the Supplementary Material. Data from the planned experimental validation (§4.7) do not yet exist and will be deposited as those experiments are completed; no planned-experiment result is reported in this version.

Repository contents (regenerable from raw data)

The following data files are tracked in the repository and are sufficient to reproduce all manuscript numbers and all quantitative figure panels:

- 02_results/global_analysis/hmm_results/*.tblout — HMMER (PF00990, PF00563, PF01966, etc.) domain-scan output for all 51 genomes (source of GGDEF/EAL/HD-GYP/PilZ domain counts)
- 02_results/global_analysis/global_signal_summary_51_corrected.tsv — per-genome signal-domain count matrix (the canonical summary used throughout the manuscript)
- 02_results/global_analysis/downloaded_genomes.tsv — genome accession list with assembly metadata
- 02_results/global_analysis/genome_environment_audit_51.json — NCBI BioSample habitat/assembly-metadata audit for all 51 genomes underlying the §4.1 habitat composition and quality assessment (isolation source, geographic location, assembly size, contig count, GC%)
- 02_results/global_analysis/proteins/*.faa — annotated proteome FASTA files for all 51 genomes (43.9 MB total; used for motif extraction)
- 02_results/kryptonia_metal_pfam_hits.tsv — metalloprotein Pfam inventory (structural metalloproteins: hemes, Fe–S clusters, zinc ribbons)
- 02_results/ggdef_structures/analysis/ggdef_structure_comparison_6.json — derived structural metrics (Ca–Ca distances between catalytic-motif residues, motif residue identities, mean pLDDT per domain, aligned-residue counts)

- 02_results/ggdef_structures/analysis/ggdef_isite_electrostatics_analysis.json — derived surface and I-site metrics (surface acidic/basic residue fractions, net charge, acidic-patch counts and largest-cluster sizes, all RxxD motif positions and their distances to the A-site). Note: canonical I-site status is determined by applying the -15 to -5 residue window criterion described in Methods to the RxxD positions stored in this JSON.
- 02_results/ggdef_structures/analysis/deeptmhmm_results.txt and tmhmm2_results.txt — transmembrane topology predictions from DeepTMHMM and TMHMM (DTU Health Tech), used to determine the subcellular side on which the acidic patches reside.
- 02_results/metal_motif_scan/global_metal_motif_summary.json — output of the genome-wide metal-binding motif scan (14 motif classes across 106,651 proteins in 51 genomes): per-motif protein counts, total hit counts, genome coverage, expected vs. observed ligand counts, and ligand-spacing statistics.
- 02_results/global_analysis/proteome_acidic_background.json — whole-proteome acidic-residue (D+E) frequency across all 51 genomes (36,175,733 amino acids): D = 5.18%, E = 7.20%, D+E = 12.38%. Source of the 12.4% proteome background cited for DENLIG/EDNLIG acidic-enrichment ratios.
- 02_results/global_analysis/structural_metal_pfam_inventory.json — reproducible PFAM HMMER re-scan of structural metalloprotein families (43 PFAM profiles, E-value $\leq 1e-5$, 51 genomes): per-category protein counts (c-type cytochromes, zinc ribbons, Rieske, [4Fe-4S], heme non-c, copper, Fe-S) and total non-redundant metalloprotein count.
- 02_results/global_analysis/pfam_nucleotide_recovery_inventory.json — reproducible PFAM HMMER re-scan of nucleotide-recovery enzymes and nucleoside transporters (8 PFAM profiles: PF00149, PF02872, PF02779, PF00334, PF00478, PF06812, PF01935, PF01547; E-value $\leq 1e-5$, 51 genomes): per-domain protein counts for the c-di-GMP degradation–recovery chain (calcineurin-like PDE, 5'-nucleotidase, gmk, ndk, guaB) and nucleoside-transport proteins (CNT/NupC, Nucleoside_tra2, ABC substrate-binding).
- 02_results/metal_motif_scan/per_protein_motif_counts.json — per-protein, per-genome metal-motif occurrence counts derived from the scan above.
- 02_results/metal_motif_scan/coordination_geometry_3d.json — three-dimensional coordination geometry comparison: heme-Fe ligand distances and angles in the Kryptonite multi-heme cytochrome c AF3 model vs. carboxylate-oxygen cluster geometry in the DENLIG GGDEF domain (used to demonstrate that the acidic patches lack metal-binding geometry).
- 02_results/metal_motif_scan/hdgyr_true_coordination_geometry.json — HD-GYP active-site metal coordination geometry for the 2 genuine Kryptonite HD-GYP proteins (MCS7229639 and MGC8595393, both carrying the canonical HD-GYP[DE] catalytic signature: HD motif followed by GYPE at 25-residue spacing): pairwise distances between the catalytic His imidazole ND1 nitrogen and the adjacent Asp carboxylate O (OD1/OD2) are 5.2–7.6 Å, ideal for sandwiching one or two metal ions in the bimetallic Fe/Mn center; multiple downstream Glu residues complete the mixed His/Asp/Glu ligand set. Compared with the DENLIG GGDEF catalytic-site carboxylate arrangement (no His ligand, O atoms too dispersed).
- 02_results/metal_motif_scan/pocket_geometry_hdgyr_vs_ggdef.json — pocket geometry comparison: atom packing density (atoms within 8 Å; DENLIG 0.0499 vs. HD-GYP 0.0317–0.0504 atoms/Å³), convex-hull pocket volume (6 Å; DENLIG 373.1 vs. HD-GYP 185.3–432.4 Å³), and pocket openness (fraction of 200 random directions from centroid reaching solvent within 6 Å; DENLIG 0.025 vs. HD-GYP mean 0.153) for HD-GYP (2 proteins: MCS7229639 and MGC8595393, both carrying canonical HD-GYP[DE] signature) vs. GGDEF DENLIG. This confirms that the GGDEF metal-binding deficiency is a ligand-composition/geometry defect (no His ligands, dispersed carboxylate O), not a solvent-exposure or pocket-shape defect.
- 02_results/ggdef_6variant_reproduction_detail.json — reproduction verification output (12/12 checks passed)
- 02_results/dnds_analysis/dnds_results.json — all dN/dS results underlying §2.10: per-family M0 ω and lnL, $\omega = 1$ likelihood-ratio tests, M7/M8 LRTs (p-values, pi), M8 Bayes Empirical Bayes positive sites, pairwise dN/dS distributions (medians, IQRs, numbers of informative pairs), Mann–Whitney *U* tests, and the GGDEF A-site/core/acidic-patch partition M0 results with their $\omega = 1$ LRTs, including alignment-based mapping of BEB sites and patch residues
- 02_results/dnds_analysis/{candidates.json, sets_manifest.json, sets/, alignments/, codeml/} — candidate proteins and motif annotations, the ortholog-cluster manifests (genome memberships, identity-dereplication maps), per-set protein and CDS FASTA files, alignments, guide trees, sequence-ID maps, and raw codeml outputs. The genomic CDS files used for back-translation were downloaded from NCBI.
- 02_results/dnds_analysis/hyphy_summary.json and 02_results/dnds_analysis/hyphy/{ggdef_fossil, hdgyr_core, hk_recA, hk_gyrB, hk_rpoB, hk_dnaG}/{busted, fel, relax}.json (+ absrel.json for ggdef_fossil) — HyPhy outputs underlying the §2.10 HyPhy corroboration paragraph: BUSTED p-values (gene-wide episodic positive selection), RELAX K and p-values (relaxation/intensification parameter), FEL per-site dN/dS (including the two M8 BEB sites 180 and 475), and aBSREL per-branch episodic selection. Reproducible with the per-set codon.fasta + tree.nwk inputs already in 02_results/dnds_analysis/alignments/.
- 02_results/cross_phylum/cross_phylum_signal_counts.json, 02_results/cross_phylum/ref_proteomes/*.faa, and 02_results/cross_phylum/hmm_results/*.tblout — HMMER-verified cross-phylum signal-domain comparison underlying §2.11: unique-protein counts per signal-domain HMM per organism (*Aquifex aeolicus* VF5, *Thermus thermophilus* HB8, *E. coli* K12, *Pyrococcus* sp., and Kryptonite 51 genomes pooled) using the same signal-domain HMM database as for Kryptonite.
- 02_results/figures/*.png and *.pdf — rendered figure panels (Figure 1, Figure 2, Figure S1–S4)

Files NOT in the repository (deposited at Zenodo, DOI 10.5281/zenodo.23015087)

The AlphaFold3 raw prediction outputs are excluded from the repository due to file size and are deposited at Zenodo. These files, organized under 02_results/ggdef_structures/<protein>/, are:

- *_model.cif — mmCIF coordinate files for the six predicted GGDEF structures (one ranked model per protein; the per-seed seed-*/ directories also contain *_model.cif for all 30 seeds)
- *_data.json — AlphaFold3 input job specification (query sequences and model seeds)
- *_confidences.json and *_summary_confidences.json — per-seed confidence data (atom-level pLDDT and PAE arrays)
- *_ranking_scores.csv — model ranking scores

Additional AlphaFold3 structures used for the 3D coordination-geometry comparison are also deposited at Zenodo:

- 02_results/af3_structures/WP143_4heme_af3_hec_bonded_*/WP143_4heme_af3_hec_bonded_model.cif — multi-heme cytochrome c structure (source of the genuine heme-Fe coordination geometry used as the reference for comparison with GGDEF acidic-patch carboxylate geometry)
- 02_results/hdgyp_structures/<protein>/_model.cif — HD-GYP phosphodiesterase structures

These files are required only to regenerate the coordinate-dependent figure panels (see below) and to rebuild the derived JSONs; all quantitative claims in the manuscript can be verified from the repository data alone.

Coordinate-file dependency for figure rendering

The following figure panels require the mmCIF coordinate files (deposited at Zenodo) and cannot be rendered from repository data alone:

- **Figure S2, panel A** — α -backbone cartoon superposition of the six GGDEF domains (requires mmCIF files for structural alignment and cartoon rendering)
- **Figure S3, panel B** — surface electrostatic potential maps (requires mmCIF files for solvent-accessible-surface calculation and Coulomb coloring)

All other quantitative panels (Figure 1, Figure 2, Figure S1, Figure S3 panels A/C/D/E/F/G, Figure S4) are regenerable from repository data alone (HMMER .tblout files, global_signal_summary_51_corrected.tsv, kryptonite_metal_pfam_hits.tsv, global_metal_motif_summary.json, and the analysis JSONs above). The quantitative metrics underlying the coordinate-dependent panels— α - α distances 3.81–3.91 Å, mean pLDDT 65–75, acidic-surface fractions (DENLIG 24.3%, GGEEF 19.3%), acidic-patch counts (DENLIG ≥ 5), heme-Fe ligand distances (1.5–2.1 Å) and ligand-Fe-ligand angles (88–92°/177–180°), HD-GYP HD-motif His N-Asp O distances (5.2–7.6 Å), DENLIG carboxylate-O centroid distances (1.76–3.18 Å), and pocket geometry metrics (DENLIG openness 0.025 vs. HD-GYP mean 0.153; DENLIG atom density 0.0499 vs. HD-GYP 0.0317–0.0504 atoms/Å³)—are verified independently from the included JSONs and do not require the coordinate files. Regeneration of the mmCIF coordinate files requires rebuilding the Jackhmmer MSAs (the full MSA files are not bundled due to size) and re-running AlphaFold3 with the deposited *_data.json inputs.

Author Contributions

Han Qinan: conceptualization, data curation, formal analysis, investigation, writing—original draft, writing—review & editing. **Zhang Minxin:** supervision, writing—review & editing.

Competing Interests

The authors declare that they have no competing financial interests. [To be updated at formal submission stage if patent applications are filed.]

Acknowledgments

We thank the hot-spring sampling teams for sample collection and the sequencing facilities for data generation. This work received no external funding.

Figures

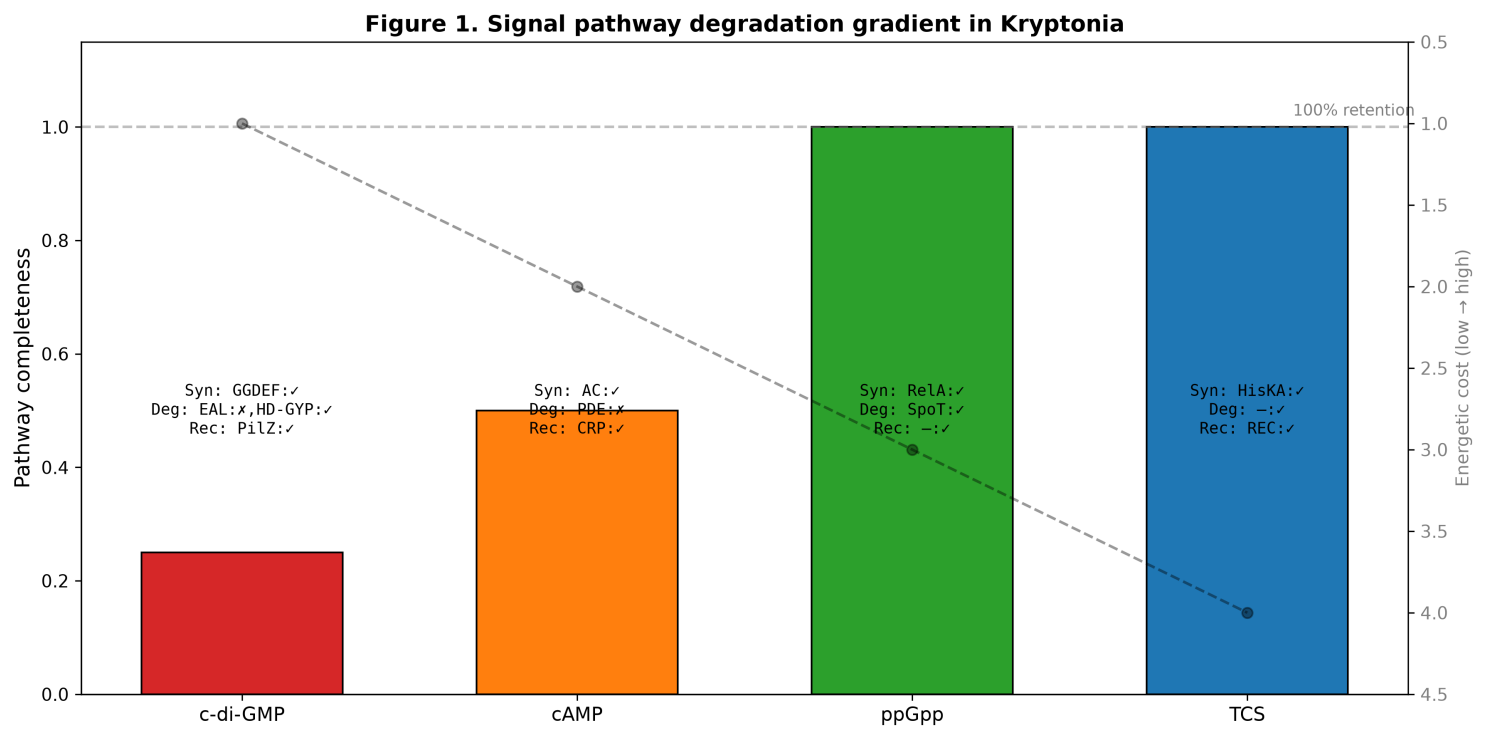


Figure 1

Figure 1. Signal pathway degradation gradient in Kryptonia. Bars represent pathway completeness (left y-axis, 0–1.0) across four bacterial signal transduction systems: c-di-GMP, cAMP, ppGpp, and two-component system (TCS). The dashed line with filled circles indicates relative energetic cost of each pathway (right y-axis, low to high, scaled 1.0–4.5). The horizontal dashed line at 1.0 denotes 100% retention. For each pathway, annotated enzyme classes include: Sym, synthase/synthetase; Deg, degradation enzymes; Rec, receptor. c-di-GMP pathway shows the highest degree of degradation (completeness $\approx$ 0.25) with all three components (CODF synthase, EAL/HD-GYP phosphodiesterases, PilZ receptor) present but partially degraded; cAMP pathway shows partial retention (completeness $\approx$ 0.50) with AC, PDE, and CRP; ppGpp and TCS pathways are fully retained (completeness = 1.0), with RelA/SpdT and HiskK/RR, respectively. The inverse relationship between pathway completeness and energetic cost indicates that Kryptonia has selectively degraded low-cost signaling systems while retaining high-cost ones during genome reduction.

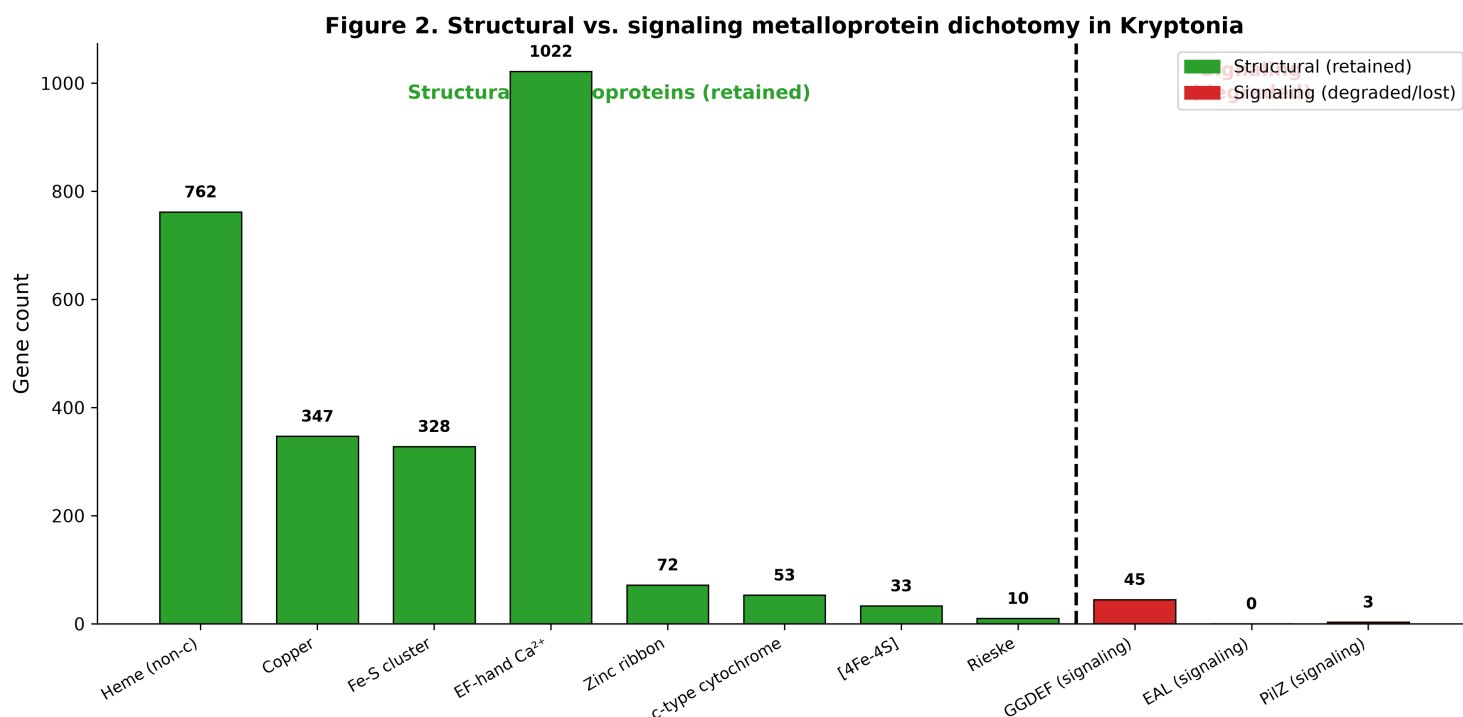


Figure 2

Figure 2. Structural versus signaling metalloprotein dichotomy in Kryptonia. Bar chart shows gene counts (y-axis) of metalloprotein families annotated from Kryptonia metagenome-assembled genomes. Green bars indicate structurally essential metalloproteins that are retained, including hemoglobin-type c (762), copper-binding proteins (347), Fe-S cluster proteins (328), EF-hand Ca²⁺-binding proteins (1022, the most abundant), zinc-ribbon proteins (72), H-type cytochromes (53), [4Fe-4S] cluster proteins (33), and Rnfase (10). Red bars indicate signaling-associated metalloproteins that are largely degraded or lost: DGC/EAL (45), EAL (0), and PilZ (3). The vertical dashed line separates structural metalloproteins (left, retained) from signaling metalloproteins (right, degraded/lost). The stark contrast—retention of hundreds to over a thousand structural metalloprotein genes versus near-complete loss of signaling metalloprotein genes—supports a genome-reduction strategy in which Kryptonia maintains core structural metal biochemistry while discarding expendable signaling pathways.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [figureS1ggdefmotifconservation.png](#)
- [FigureS2GGDEFstructurecomparison.png](#)
- [FigureS3isiteelectrostatics.png](#)
- [figureS4helpercheater.png](#)