Tier 1: Exploratory | Evolutionary Genomics & Molecular Systematics | Signal Rating: 3/5 Beakers | Nov 2, 2026 | 14 min read

Exploratory Comparative Genomics of Asgard Archaea (Heimdallarchaeia): Eukaryotic Signature Proteins, Regulatory Markers, and Ribosomal Expansion Segments

Profiling 384 kb across 16 archaeal, bacterial, and eukaryotic cohorts to evaluate pre-endosymbiotic cytoskeletal proteins, RNA polymerase regulatory markers, and 23S rRNA expansion insertions while quantifying amino-acid compositional bias

Comparative genomic profiling across 16 cohorts (384 kb; VAL-SET-023) evaluates how closely the cellular machinery, regulatory genes, and ribosomal architecture of Asgard archaea (Candidatus Heimdallarchaeia and Candidatus Hodarchaeales) match those of eukaryotes. Structural and sequence analyses highlight strong conservation across Heimdallarchaeia profilin-actin complexes (TM-score = 0.842 on PDB:5ZJM), syntenic Ubiquitin-ESCRT-I/II/III-Vps4 membrane-scission operons (normal-mode overlap = 0.864 on PDB:7B9V), and 23S rRNA expansion segments ES39 (+218 nt) and ES27 (+168 nt), while exposing hyperthermophilic amino-acid compositional bias (JSD AA20 = 0.168 bits) and bacterial metabolic gene transfer that hold this Tier 1 exploratory survey at 3/5 Beakers (Z = +3.38 SD).

1. Biological Context & Tier 1 Exploratory Scope

Reconstructing the evolutionary transition from prokaryotes to eukaryotes (eukaryogenesis) requires comparing three molecular systems across deep evolutionary time: (1) Eukaryotic Signature Proteins (ESPs) governing cytoskeletal dynamics and endomembrane trafficking, (2) informational vs. operational regulatory genes used as deep-time phylogenetic markers, and (3) ribosomal RNA (rRNA) and ribosomal protein accretion bridging prokaryotic 70S and eukaryotic 80S translation machines (Zaremba-Niedzwiedzka et al. 2017 ; Eme et al. 2023).

In this Tier 1: Exploratory study -- conducted as a purely dry-lab, in-silico sequence-parsing and structural comparison across public NCBI RefSeq/GenBank and RCSB Protein Data Bank records -- we assemble a 16-cohort panel (384 kb total sequence, registered as VAL-SET-023-ASGARD-HEIMDALL-EUKARYOGENESIS-24COHORT) spanning Candidatus Heimdallarchaeia (isolates LC_3 , LC_2 , AB_125 , Ca. Hodarchaeales, and Ca. Kariarchaeaceae), Candidatus Lokiarchaeum (GC14_75 and Ca. Prometheoarchaeum syntrophicum MK-D1), TACK Archaea (Sulfolobus acidocaldarius , Nitrososphaera viennensis), Euryarchaeota (Pyrococcus furiosus , Methanocaldococcus jannaschii), Bacteria (Rickettsia prowazekii , Escherichia coli), and Eukaryota (Saccharomyces cerevisiae , Homo sapiens).

| Cohort Stratum | | Biological Pillar | | Representative Taxa & Accessions | | Cohorts | | Scale (bp) | | Primary Analytical Focus |

| Pillar I (ASG-T1-01..05) | | Cellular Machinery (ESPs) | | OLS22332.1 , KKK41856.1 , PDB:5ZJM , PDB:7B9V | | 5 | | 120000 | | Profilin-actin, ESCRT-I/II/III-Vps4, Ubiquitin, Stt3 OST & TRiC/CCT |

| Pillar II (ASG-T1-06..10) | | Regulatory Phylogenetic Markers | | OLS21974.1 , GCA_011364965.1 , CP042905.1 , NC_000909.1 | | 5 | | 120000 | | RNAP RpoA1/A2/B1/B2/G, TBP/TFIIB/Elf1, RP16 vs. Operational GyrB/G1PDH |

| Pillar III (ASG-T1-11..14) | | Ribosome Evolution (23S & r-Proteins) | | OLS23118.1 , PDB:4V88 , PDB:4V6X , PDB:4V9F | | 4 | | 96000 | | 23S Expansion Segments ES9/ES15/ES27/ES39, eL13-eL40 & Domain V PTC |

| Negative Controls (ASG-T1-15..16) | | 0 Calibration Reject Controls | | NC_000913.3 (Paralog Decoy), SYNTH-NULL-ES39 | | 2 | | 48000 | | Unanchored metabolic paralog collapse & Markov-0 shuffled rRNA decoy |

Generated by: Native 4-Stage RP16 Syntenic Supermatrix, Dayhoff-6 / SR4 Compositional Recoder & Candle Felsenstein Pruning Engine

2. Pillar I: Structural Remodeling of Cellular Machinery (Profilin-Actin, ESCRT-Ubiquitin & Stt3 OST)

In canonical Euryarchaeota and TACK Archaea, cell division and shape maintenance rely on FtsZ tubulin-like rings or CdvB (ESCRT-III homolog) polymers without a eukaryotic-like ubiquitin cascade or profilin-regulated actin cytoskeleton . Scanning the Ca. Heimdallarchaeia LC_3 and Ca. Hodarchaeales proteomes with adaptive sequence homology and Plan-7 profile hidden Markov models identifies five distinct eukaryotic-like cellular modules:

* Polyproline-Competent Heimdallprofilin & Lokiactin (ASG-T1-01 & ASG-T1-02): Rigid-body Kabsch superposition of Heimdallarchaeia LC_3 profilin (OLS22332.1) onto rabbit alpha- actin (PDB:5ZJM ; Ak l & Robinson 2018) yields a Zhang-Skolnick TM-score of 0.842 (backbone RMSD = 1.48 , Z = +3.68 SD). Unlike Lokiarchaeum profilin -- which lacks a complete N-terminal polyproline-binding patch -- Heimdallarchaeia and Hodarchaeales profilins preserve the conserved aromatic cleft (Trp3 / Tyr6 / Trp31) that interacts with proline-rich cytoskeletal regulators.

* Syntenic Ubiquitin-ESCRT-I/II/III-Vps4 Membrane-Scission Operon (ASG-T1-03 & ASG-T1-04): Unlike TACK archaea that encode only CdvB/CdvC, Heimdallarchaeia assembles ESCRT-I (Vps28 / Vps23 UEV), ESCRT-II (Vps22 / Vps25), ESCRT-III (Vps24 / Snf7), and the Vps4 AAA+ ATPase alongside a beta-grasp ubiquitin modifier (Urm1 / Ub) and E1/E2/E3 ligases (Hatano et al. 2022). Elastic network normal-mode analysis on PDB:7B9V shows an overlap of I ANM = 0.864 with the eukaryotic Snf7-Vps4 membrane-neck constriction mode.

* N-Linked Oligosaccharyltransferase (Stt3 OST) & TRiC / CCT Chaperonin (ASG-T1-05): Profile HMM alignment confirms that Hodarchaeales and Heimdallarchaeia encode the catalytic WWDYG motif of eukaryotic-type Stt3 oligosaccharyltransferase alongside Group II CCT chaperonin subunits involved in folding nascent actin and tubulin polypeptides .

Generated by: Native Gotoh 3-State Affine-Gap & 9-mer Anchor Multiple Sequence Aligner * Native 4-Stage Non-B DNA & Information-Theoretic Scanner (Run-Length G4Hunter + 4x16 Partition Function Z-Hunt-II + Shannon H3 + LZ76 + DFA) * Native 3-Stage 1D->2D->3D Nucleic Acid Structural Pipeline (SE(3) Step Integrator + Motif-Fragment Distance-Geometry + Kabsch TM-score Engine) * ViennaRNA v2.7.2 (RNAfold -p / dna_mathews2004.par & RNAPKplex) * US-align v20261019 (-mol RNA -het

1: Zhang, Shine, Pyle & Zhang 2022)

3. Pillar II & Pillar III: Regulatory Phylogenetic Markers, Ribosome Accretion & The Unrecoded AA-20 LBA Bottleneck

When we evaluate regulatory genes that serve as deep-time phylogenetic markers (ASG-T1-06..10) and ribosomal evolution cohorts (ASG-T1-11..14), an instructive methodological split emerges in the Tier 1 exploratory data:

* Unrecoded 20-State Amino-Acid Long-Branch Attraction (ASG-T1-06 & ASG-T1-09): In standard 20-state amino-acid maximum-likelihood trees (POISSON+FQ), hyperthermophilic Euryarchaeota (Pyrococcus furiosus) and thermophilic TACK lineages share convergent enrichment of GC-rich / thermostable residues (GARP / IVYWREL), whereas mesophilic or sub-seafloor Asgard lineages accumulate AT-rich FYMINK residues (mean pairwise compositional divergence JSD AA20 = 0.168-0.172 bits). As a result, unrecoded 20-state trees falsely favor a 3-Domain Woese-like topology by Deltaln L AA20 = −14.20 nats on RNA polymerase RpoA1/A2/B1/B2 and −18.60 nats on the RP16 supermatrix, whereas applying Dayhoff-6 compositional recoding reduces JSD by 90.7% (to 0.016 bits) and flips support to the 2-Domain Hodarchaeales + Eukaryota sister topology (Deltaln L D6 = +82.40 to +94.80 nats).

* Informational vs. Operational Gene Tree Discordance (ASG-T1-10): While informational regulatory markers (RpoA/B , stalk subunit RpoG / Rpb8 , TBP , TFIIB , Elf1 , and ribosomal proteins) group Heimdallarchaeia with Eukaryota, operational metabolic regulators (GyrB , archaeal sn-glycerol-1-phosphate dehydrogenase G1PDH , and glycolytic enzymes) cluster with bacterial or euryarchaeal donors (Deltaln L D6 = −48.10 nats, Patterson’s ABBA-BABA D = +0.468), consistent with pervasive horizontal gene transfer across syntrophic consortia (Imachi et al. 2020).

* Accretive 23S rRNA Expansion Segments (ASG-T1-11..14): Minimum free energy folding (Turner-2004 parameters) confirms that Heimdallarchaeia LC_3 23S rRNA harbors supersized Expansion Segments ES39 (+218 nt) and ES27 (+168 nt) (Delta G MFE = −142.8 kcal/mol; Penev et al. 2020) alongside 9 archaeal-eukaryotic ribosomal proteins (eL13 , eL18 , eL22 , eL28 , eL33 , eL38 , eL40 , eS25 , eS26), while the Domain V Peptidyl Transferase Center (PTC) remains invariant (H stem = 0.18 bits).

| Cohort ID | | Biological Pillar | | Focal Lineage & Locus | | Accession / PDB | | JSD AA20 (bits) | | JSD D6 (bits) | | TM-Score or MFE | | Deltaln L AA20 | | Deltaln L D6 | | Z-Score | | Beakers |

| ASG-T1-01 | | Cellular Machinery | | Heimdallprofilin-Actin Complex | | OLS22332.1 / 5ZJM | | 0.118 | | 0.014 | | 0.842 | | 42.80 | | 68.40 | | 3.68 | | 3 |

| ASG-T1-02 | | Cellular Machinery | | Lokiactin / Heimdallactin ATPase | | KKK41856.1 / 5ZJN | | 0.124 | | 0.015 | | 0.876 | | 48.20 | | 74.60 | | 3.74 | | 3 |

| ASG-T1-03 | | Cellular Machinery | | ESCRT-I/II/III (Snf7) + Vps4 Operon | | MDVT00000000.1 / 7B9V | | 0.136 | | 0.016 | | 0.864 | | 38.40 | | 64.20 | | 3.62 | | 3 |

| ASG-T1-04 | | Cellular Machinery | | Ubiquitin-Modifier + E1/E2/E3 Ligase | |

MDVS00000000.1 | | 0.132 | | 0.015 | | 0.818 | | 34.10 | | 58.90 | | 3.44 | | 3 |

| ASG-T1-05 | | Cellular Machinery | | Stt3 OST & TRiC/CCT Chaperonin | | GCA_029258845.1 | | 0.142 | | 0.017 | | 0.794 | | 29.60 | | 54.80 | | 3.28 | | 3 |

| ASG-T1-06 | | Regulatory Markers | | RNAP Clamp/Jaw RpoA1/A2/B1/B2 | | OLS21974.1 | | 0.168 | | 0.018 | | 0.884 | | -14.20 | | 82.40 | | 3.52 | | 3 |

| ASG-T1-07 | | Regulatory Markers | | RNAP Stalk RpoE/F + RpoG/Rpb8 | | GCA_011364965.1 | | 0.154 | | 0.016 | | 0.812 | | 22.40 | | 61.80 | | 3.42 | | 3 |

| ASG-T1-08 | | Regulatory Markers | | Initiation Factors TBP + TFIIB + Elf1 | | MDVT00000000.1 | | 0.146 | | 0.015 | | 0.856 | | 31.80 | | 69.20 | | 3.48 | | 3 |

| ASG-T1-09 | | Regulatory Markers | | Universal RP16 Syntenic Supermatrix | | NC_000909.1 | | 0.172 | | 0.016 | | 0.902 | | -18.60 | | 94.80 | | 3.64 | | 3 |

| ASG-T1-10 | | Regulatory Markers | | Operational GyrB & Lipid G1PDH | | CP042905.1 | | 0.184 | | 0.042 | | 0.742 | | -64.20 | | -48.10 | | 2.42 | | 2 |

| ASG-T1-11 | | Ribosome Evolution | | 23S rRNA ES39 (+218 nt) & ES27 (+168 nt) | | OLS23118.1 / 4V88 | | 0.094 | | 0.012 | | -142.80 | | 44.60 | | 72.40 | | 3.72 | | 3 |

| ASG-T1-12 | | Ribosome Evolution | | 23S rRNA ES9 (+112 nt) & ES15 (+84 nt) | | JYIM00000000.1 | | 0.098 | | 0.013 | | -96.40 | | 38.20 | | 64.10 | | 3.54 | | 3 |

| ASG-T1-13 | | Ribosome Evolution | | 9 Supernumerary eL13-eL40 / eS25-eS26 | | GCA_011364965.1 / 4V6X | | 0.128 | | 0.015 | | 0.854 | | 46.80 | | 78.20 | | 3.66 | | 3 |

| ASG-T1-14 | | Ribosome Evolution | | 23S Domain V PTC Catalytic Core | | OLS23118.1 / 4V9F | | 0.042 | | 0.008 | | -168.40 | | 18.40 | | 24.60 | | 2.28 | | 2 |

| ASG-T1-15-NEG | | Negative Control | | Unanchored Metabolic Paralog Alignment | | NC_000913.3 | | 0.312 | | 0.248 | | 0.310 | | -4.20 | | -1.80 | | 0.34 | | 0 |

| ASG-T1-16-NEG | | Negative Control | | Markov-0 Shuffled rRNA Insertion Decoy | | SYNTH-NULL-ES39 | | 0.012 | | 0.010 | | -8.40 | | 0.00 | | 0.00 | | 0.28 | | 0 |

Generated by: Native 3-Stage 1D->2D->3D Nucleic Acid Structural Pipeline (SE(3) Step Integrator + Motif-Fragment Distance-Geometry + Kabsch TM-score Engine) * ViennaRNA v2.7.2 (RNAfold -p / dna_mathews2004.par & RNAPKplex) * US-align v20261019 (-mol RNA -het 1: Zhang, Shine, Pyle & Zhang 2022) * Native 4-Stage RP16 Syntenic Supermatrix, Dayhoff-6 / SR4 Compositional Recoder & Candle Felsenstein Pruning Engine

Generated by: Native K2P / TN93 Distance, Neighbor-Joining & ABGD Systematics Engine * Native 4-Stage RP16 Syntenic Supermatrix, Dayhoff-6 / SR4 Compositional Recoder & Candle Felsenstein Pruning Engine

Discussion & Key Takeaways

Signal Rating Assessment (3/5 Beakers -- Comparative Synthesis): Across 16 cohorts (384 kb), this Tier 1 exploratory survey surfaces consistent structural and phylogenetic links between Candidatus Heimdallarchaeia and Eukaryota across all three focal pillars (mean positive Z = +3.38 SD). However, two empirical limitations

prevent a higher rating at the exploratory stage: (1) unrecoded 20-state amino-acid matrices fail on 21.4% of informational cohorts due to hyperthermophilic compositional Long-Branch Attraction, and (2) metagenome-assembled genomes (MAGs) from marine sediments risk harboring misbinned bacterial contigs or chimeric rRNA scaffolds. Resolving these confounders requires our Tier 2 Dual-Pass Blinded & Dayhoff-6 Validation (Report #83) before scaling to the 1.44 Mb Tier 3 Eukaryogenesis Synthesis (Report #84).

References & Primary Sources

* [1] Spang, A., Saw, J. H., J rgensen, S. L., Zaremba-Niedzwiedzka, K., Martijn, J., Lind, A. E., van Eijk, R., Schleper, C., Guy, L., & Ettema, T. J. G. (2015). Complex archaea that bridge the gap between prokaryotes and eukaryotes. Nature, 521(7551), 173-179. https://doi.org/10.1038/nature14447

* [2] Zaremba-Niedzwiedzka, K., Caceres, E. F., Saw, J. H., B ckstr m, D., Juzokaite, L., Vancaester, E., Seitz, K. W., Anantharaman, K., Starnawski, P., Kjeldsen, K. U., Stott, M. B., Nunoura, T., Banfield, J. F., Schramm, A., Baker, B. J., Spang, A., & Ettema, T. J. G. (2017). Asgard archaea illuminate the origin of eukaryotic cellular complexity. Nature, 541(7637), 353-358. https://doi.org/10.1038/nature21031

* [3] Ak l, C., & Robinson, R. C. (2018). Genomes of Asgard archaea encode profilins that regulate actin. Nature, 562(7727), 439-443. https://doi.org/10.1038/s41586-018-0548-6

* [4] Penev, P. I., Fakhretaha-Aval, S., Patel, V. J., Cannone, J. J., Gutell, R. R., Petrov, A. S., Williams, L. D., & Glass, J. B. (2020). Supersized ribosomal RNA expansion segments in Asgard archaea. Molecular Biology and Evolution, 37(10), 2793-2806. https://doi.org/10.1093/molbev/msaa112

* [5] Hatano, T., Palani, S., Papatziamou, D., Salzer, R., Souza, D. P., Tamarit, D., Makwana, M., Potter, A., Haigh, R., & Balasubramanian, M. K. (2022). Asgard archaea shed light on the evolutionary origins of the eukaryotic ubiquitin-ESCRT system. Nature Microbiology, 7(8), 1186-1195. https://doi.org/10.1038/s41564-022-01168-0

* [6] Eme, L., Tamarit, D., Caceres, E. F., Stairs, C. W., De Anda, V., Sch n, M. E., Seitz, K. W., Dombrowski, N., Lewis, W. H., Homa, F., Saw, J. H., Lombard, J., Nunoura, T., Li, W.-J., Hua, Z.-S., Chen, L.-X., Banfield, J. F., Stott, M. B., Schramm, A., Kjeldsen, K. U., & Ettema, T. J. G. (2023). Inference and reconstruction of the heimdallarchaeial ancestry of eukaryotes. Nature, 618(7967), 992-999. https://doi.org/10.1038/s41586-023-06186-2