

Table S1: tRNA-Gly^{GCC} of 70 Archaea, whose number of hexamers CTGCCA observed is 37 and expected $70 \times 70 / 46 = 1.2 \pm 1.8$ (95% CI)

Candidatus Korarchaeum cryptofilum OPF8, complete sequence NCBI Reference Sequence: NC_010482.1: c1321063-1320990 tRNA-Gly

GCGGCCGTAGTCTAGCC**TGG**ACAGGATGGGGGC**CTGCCA**CGTCCCAGACCCGGG**TTCAA**ATCCCGGCGGCCGCA

Thermofilum pendens Hrk 5, complete sequence NCBI Reference S equence: NC_008698.1: 796113-796158,796179-796209 tRNA-Gly

GCGGCCGTAGTCTAGTC**TGGTA**AGGATGGCGGC**CTGCCA**CGCCGCAGAACCCGGG**TTCAA**ATCCCGGCGGCCGCACCA

Caldivirga maquilingensis tRNA-Gly (GCC), complete sequence, strain: IC-167 GenBank: AB470218.1 tRNA-Gly

GCGGCCGTAGTCTAGTC**TGGT**TTAGGATGGCGGC**CTGCC**GCGCCGCAGATCCCGGG**TTCAA**ATCCCGGCGGCCGCA

Vulcanisaeta distributa DSM 14429, complete genome GenBank: CP002100.1: 953496-953593 tRNA-Gly

GGGCCCCGTAGTCTAGCC**TGGTA**AGGATGCCCAGGGGCCGA**TGCC**CCGTACAAGCCTGACGCGCGGGAGATC
CCGGG**TTCAAG**TCCCGGCGGGCCCCACCA

Thermoproteus tenax Kra 1, complete genome GenBank: FN869859.1: c40937-40818 tRNA-Gly

GCGGCGGTAGTCTAGCC**TGGT**TAGGGCGGCCAAAAGCCCCAGAGTAGGATGGCGGC**CTGCCA**GGCCGCA
CCATCGGCGCAGTGAGCCGTTGATCCCGGG**TTCAA**ATCCCGGCCCGCCA

Pyrobaculum neutrophilum V24Sta, NCBI Reference Sequence: NC_010525.1: 1441482-1441503,1441534-1441590 tRNA-Gly

GCGGCGGTAGTCTAGCC**TGGT**TTAGGATGGCGGC**CTGCCA**AGCCGTTGATCCCGGG**TTCAA**ATCCCGGCC
GCCGCACCA

Pyrobaculum arsenaticum DSM 13514, complete sequence NCBI Reference Sequence: NC_009376.1: c1479953-1479930,c1479895-1479841 tRNA-Gly

GCGGCGGTAGTCTAGCC**TGGT**TTAGGATGGCGGC**CTGCCA**AGCCGTTGATCCCGGG**TTCAA**ATCCCGGCCGCCGCACCA

Pyrobaculum aerophilum str. IM2, complete sequence NCBI Reference Sequence: NC_003364.1: 1904041-1904119 tRNA-Gly

GCGGCGGTAGTCTAGCC**TGGT**TTAGGATGGCGGC**CTGCCA**AGCCGTTGATCCCGGG**TTCAA**ATCCCGGCCGCCGCACCA

Sulfolobus tokodaii str. 7 DNA, complete genome GenBank: BA000023.2: 1719268-1719343 tRNA-Gly

GCGGCCGTAGTCTAGCC**TGG**ATTAGGACGCCGGC**CTGCCA**CGCCGAGGTCCCGGG**TTCAA**ATCCCGGCGGTTCGCA

Sulfolobus solfataricus strain P1 genome assembly, chromosome: I GenBank: LT549890.1: 388180-388255 tRNA-Gly

GCGGCCGTAGTCTAGCC**TGG**ATTAGGACGCCTGC**CTGCCA**CGCAGGAGGTCCCGGG**TTCAA**ATCCCGGCGGCCGC

Metallosphaera yellowstonensis MK1 genomic scaffold MetMK1scaffold_19, whole genome shotgun sequence GenBank: JH597770.1: 457522-457597 tRNA-Gly

GCGGCCGTAGTCTAGCC**TGG**ATTAGGACGCCTGC**CTGCCA**CGCAGGAGGTCCCGGG**TTCAA**ATCCCGGCGGTTCGCA

Metallosphaera cuprina Ar-4, complete sequence NCBI Reference Sequence: NC_015435.1: 110857-110932 tRNA-Gly

GCGGCCGTAGTCTAGCC**TGG**ATTAGGACGCCTGC**CTGCCA**CGCAGGAGGTCCCGGG**TTCAA**ATCCCGGCGGTTCGCA

Ignicoccus hospitalis KIN4/I, complete sequence NCBI Reference Sequence: NC_009776.1: 289795-289873 tRNA-Gly

GCGGCGGTCTGTCTAGCC**TGG**ACTAGGACGCCGGC**CTGCCA**CGCCGAGATCCCGGGTTC**AA**ATCCCGGCCGCCGCACCA

Hyperthermus butylicus DSM 5456, complete sequence NCBI Reference Sequence: NC_008818.1: c376263-376186 tRNA-Gly

GCGGCGGTCTGTCTAGCC**TGG**ACTAGGACGCCGGC**CCCCA**AGCCGGTCACCCGGG**TTCAA**ATCCCGGCCGCCGCACCA

Aeropyrum pernix K1, complete sequence NCBI Reference Sequence: NC_000854.2: c467913-467835 tRNA-Gly

GCGGCGGTCTGTCTAGCC**TGG**ACTAGGACGCCGGC**CTGCCA**AGCCGGAGGTCCCGGG**TTCAA**ATCCCGGCCGCCGCACCA

Ignisphaera aggregans DSM 17230, complete genome GenBank: CP002098.1: c1027071-1026993 tRNA-Gly

GCGGCGGTCTGTCTAGCC**TGGT**CTAGGACGCCGGC**CTGCCA**AGCCGGAGATCCCGGG**TTCAA**ATCCCGGCCGCCGCACCA

Staphylothermus marinus F1, complete sequence NCBI Reference Sequence: NC_009033.1: 624851-624890,624916-624954 tRNA-Gly

GCGGCGGTCTGTCTAGCC**TGG**ACTAGGACG**CCGGC****CTGCCA**CGCCGGAGATCCCGGG**TTCAA**ATCCCGGCCGCCGCACCA

Thermosphaera aggregans strain 3507 chromosome, Reference Sequence: NZ_CP063144.1: 483817-483892 tRNA-Gly

GCGGCGGTCTGTCTAGCC**TGG**ACTAGGACGCCGGC**CTGCCA**CGCCGGAAATCCCGGG**TTCAA**ATCCCGGCCGCCGCA

Desulfurococcus mucosus DSM 2162, complete sequence NCBI Reference Sequence: NC_014961.1: c1034649-1034574 tRNA-Gly

GCGGCGGTCTGTCTAGCC**TGG**ACTAGGACGCCGGC**CTGCCA**CGCCGGAAATCCCGGG**TTCAA**ATCCCGGCCGCCGCA

Desulfurococcus fermentans DSM 16532, complete genome GenBank: CP003321.1: 47280-47355 tRNA-Gly

GCGGCGGTCTGTCTAGCC**TGG**ACTAGGACGCCGGC**CTGCCA**CGCCGGAAATCCCGGG**TTCAA**ATCCCGGCCGCCGCA

Caldiarchaeum subterraneum DNA, complete genome, from subsurface mine microbial mat GenBank: BA000048.1: c1207947-1207874 tRNA-Gly

GCGGCGGTCTGTCTAGCC**TGGTA**GGACACCAGC**CTGCCA**CGCTGGGAGTCGCGGG**TTCAA**ATCCCGCCCGCCGCA

Cenarchaeum symbiosum A, complete genome GenBank: DP000238.1: c1160652-1160579 tRNA-Gly

GCAGTCATAGTATAGTT**TGGT**TATTATTCCAGCT**TGCCA**AGTTGGTGACCCGGG**TC****CAA**ATCCCGGTGACTGCA

Nitrosopumilus maritimus SCM1, complete genome GenBank: CP000866.1: c397959-397883 tRNA-Gly

GCAGTTATAGTACAGTC**TGGT**TAGTACTCGTGCT**TGCCA**AGTACGTGACCCGGG**TTCAA**ATCCCGGTAAGTGCACCA

Pyrococcus furiosus DSM 3638, complete genome GenBank: AE009950.1: c361464-361387 tRNA-Gly

GCGGTGGTAGTCTAGCC**TGGT**CTAGGACGCCACC**CTGCCA**AGGTGGAGACCCGGG**TTCAA**ATCCCGGCCACCGCACCA

Pyrococcus sp. NA2, complete sequence NCBI Reference Sequence: NC_015474.1: 1071591-1071668 tRNA-Gly

GCGGTGGTAGTCTAGCC**TGG**CCTAGGACGCCACC**CTGCCA**AGGTGGAGACCCGGG**TTCAA**ATCCCGGCCACCGCACC

Pyrococcus abyssi GE5, complete sequence NCBI Reference Sequence: NC_000868.1: c1468795-1468718 tRNA-Gly

GCGGTGGTAGTCTAGCC**TGG**CCTAGGACGCCACC**CTGCCA**AGGTGGAGACCCGGG**TTCAA**ATCCCGGCCACCGCACCA

Thermococcus sibiricus MM 739 tRNA-Gly-GCC-1-1 gene GenBank: HG986370.1

GCGGTGGTAGTCTAGCC**TGGT**CTAGGACAGCGGC**CTGCCA**CGCCGCGGGCCCGGG**TTCAA**ATCCCGGCCA
CCGCACCA

Thermococcus kodakarensis KOD1 DNA, complete genome GenBank: AP006878.1: c1311151-1311075 tRNA-Gly

GCGGTGGTAGTCTAGCC**TGG**CTAGGACAGCGGC**CTGCCA**CGCCGCTGGCCCCGGG**TTCAA**ATCCCCGGCCAC
CGCACCA

Thermococcus gammatolerans EJ3 tRNA-Gly-GCC-1-1 gene GenBank: HG986232.1

GCGGTGGTAGTCTAGCC**TGG**CTAGGACAGCGGC**CTGCCA**CGCCGCTGGCCCCGGG**TTCAA**ATCCCCGGCCACCGCACCA

Methanopyrus kandleri AV19, complete genome GenBank: AE009439.1: c1634571-1634496 tRNA-Gly

GCGGCCGCAGTCTAGTC**TGGTA**GGACGCGGGC**CTGCC**GAGCCCGTGGCCCCGGG**TTCAA**ATCCCCGGCGGCCGCACCA

**Methanothermobacter thermautotrophicus strain WHS NODE_1_length_560865_cov_91_9211, whole genome shotgun sequence
NCBI Reference Sequence: NZ_QKOF01000006.1: c518685-518612 tRNA-Gly**

GCGGCGTTAGTCCAGCC**TGGT**TAAGACACTGGC**CTGCCA**CGCCAGCGACCCGGG**TTCAA**ATCCCCGGACGCCGCA

**Methanobrevibacter smithii strain WWM1085 Contig_1, whole genome shotgun sequence NCBI Reference Sequence:
NZ_NQLD01000012.1: c378964-378891 tRNA-Gly**

GCGGTGTTAGTCCAGCC**TGGT**TAAGACTCTAGC**CTGCCA**CGTTAGAGACCCGGG**TTCAA**ATCCCCGGACGCCGCA

**Methanocaldococcus bathoardescens strain JH146 chromosome, complete genome NCBI Reference Sequence: NZ_CP009149.1:
c640173-640098 tRNA-Gly**

GCGGCCTTGG**TG**TAGCC**TGGTA**ACACACGGGC**CTGCCA**CGCCCGGACCCCGGG**TTCAA**ATCCCCGGAGGCCGCACCA

Methanocaldococcus infernus ME, complete genome GenBank: CP002009.1: c1047961-1047886 tRNA-Gly

GCGGCCTTGG**TG**TAGCC**TGGTA**ACACACGGGC**CTGCCA**CGCCCGGACCCCGGG**TTCAA**ATCCCCGGAGGCCGCACCA

Methanocaldococcus fervens AG86, complete genome GenBank: CP001696.1:c280371-280296 tRNA-Gly

GCGGCCTTGG**TG**TAGCC**TGGTA**ACACACGGGC**CTGCCA**CGCCCGGACCCCGGG**TTCAA**ATCCCCGGAGGCCGCACCA

Methanocaldococcus jannaschii DSM 2661, complete sequence NCBI Reference Sequence: NC_000909.1: 1150145-1150220 tRNA-Gly

GCGGCCT**TGGT**GTAGCC**TGGTA**ACACACGGGC**CTGCCA**CGCCCGGACCCCGGG**TTCAA**ATCCCCGGAGGCCGCACCA

Methanococcus aeolicus Nankai-3, complete genome GenBank: CP000743.1: c846618-846546 tRNA-Gly

GCGGTTTT**GATG**TAGCC**TGGTA**TCATACGGCC**CTGCCA**CGGCCGACACCCGGG**TTCGAA**TCCCCGGAGACCGCA

**Methanococcus voltae strain DSM 14649 Ga0466840_02, NCBI Reference Sequence: NZ_JAGGMO010000002.1: c253803-253733
tRNA-Gly**

GCGGCCTTG**ATGTAGTGGTA**TCATACGGCC**CTGCCA**CGGCCGATACCCGGG**TTCGAA**TCCCCGGAGGCCGCA

Methanococcus vannielii SB gene GenBank: HG986108.1 tRNA-Gly

GCGGCTTT**GATGTAGACTGGTA**TCATACGGCC**CTGCCA**CGGCCGACACCCGGG**TTCAA**ATCCCGGAGGCCGCA

Methanococcus maripaludis strain S2, complete sequence GenBank: BX950229.1: c1269120-1269048 tRNA-Gly

GCGGCTTTG**ATGTAGACTGGTA**TCATACGGCC**CTGCCA**CGGCCGACACCCGGG**TTCAA**ATCCCCGGAGGCCGCA

Aciduliprofundum boonei T469, complete genome GenBank: CP001941.1: c934319-934247 tRNA-Gly

GCGGGTG**TGGT**GTAGCC**TGGTA**ACACGAGGCC**CTGCCA**CGGCTTTGCCCCGGG**TTCAA**ATCCCCGGCACCCGCA

Thermoplasma volcanium GSS1, complete sequence NCBI Reference Sequence: NC_002689.2: 551951-552023 tRNA-Gly

GCGGGTGTGGTGTAGCC**TGG**CAACACGCGAGCT**TGCCA**AGCTCGTGCCTCGGG**TTCAA**ATCCCGACATCCGCA

Ferroglobus placidus DSM 10642, complete genome GenBank: CP001899.1: 566151-566224 tRNA-Gly

GCGCCGGTGGTGTAGCC**TGG**CTAACACAGGGGCT**TGCC**GAGCCCCTGCCCCGGG**TTCAA**ATCCCGGCCGGCGCA

Archaeoglobus profundus DSM 5631, complete sequence NCBI Reference Sequence: NC_013741.1: 1365143-1365216 tRNA-Gly

GCGCCGGTGGTGTAGCC**TGG**CTAACACAGGGGC**CTGCCA**CGCCCCAGCCCCGGG**TTCAA**ATCCCGGCCGGCGCA

Archaeoglobus fulgidus isolate Archaeoglobus_fulgidus_oil, whole genome shotgun sequence NCBI Reference Sequence: NZ_FJNF01000147.1: 2080-2152 tRNA-Gly

GCGCCG**AT**GGTGTAGCC**TGGTA**ACACAGGGGCT**TGCC**GAGCCCCT**GC**CCCCGGG**TTCAA**ATCCCGGTCTGGCGCA

Methanocella paludicola SANA E, complete sequence NCBI Reference Sequence: NC_013665.1: 2598124-2598195 tRNA-Gly

GCGCCGGTAGTGT**AGTGGT**TATCACTATGGCT**TGCCA**AGCCATAAACTCGGG**TTCGA**GTCCCGACCGGCGCA

Methanosaeta harundinacea 6Ac, complete sequence NCBI Reference Sequence: NC_017527.1: c235976-235905 tRNA-Gly

GCGCCGATAGTGT**AGTGGT**CATCACTAGGCGT**TGCCA**ACGCCTAAACCCGGG**TTCGA**ATCCCGGTCTGGCGCA

Methanosarcina barkeri CM1, complete genome GenBank: CP008746.1: 1891060-1891131 tRNA-Gly

ACATCAGTAGTGT**AGCGGT**CATCACCGGGCGT**TGCCA**ACGCTCGAACCCGGG**TTCGA**ATCCCGGTCTGATGTA

Methanosarcina mazei TMA DNA, complete genome GenBank: AP019780.1: 633659-633730 tRNA-Gly

ACATCAGTAGTGTAGC**GGT**CATCACCGGGCGT**TGCCA**ACGCTCGAACCCGGG**TTCGA**ATCCCGGTCTGGTGTGTA

Methanococcoides burtonii DSM 6242, complete genome GenBank: CP000300.1: c113010-112939 tRNA-Gly

ACACCAGTAGTGTAGC**GGT**TATCACCGGGCGT**TGCCA**ACGCTCGAACTCGGG**TTCGA**TTCCCGACTGGTGTGTA

Methanocorpusculum labreanum Z, complete sequence NCBI Reference Sequence: NC_008942.1: 337952-338023 tRNA-Gly

GCATCGATAGTGT**AGTGGT**TATCACTAGGCGT**TGCCA**ACGCCTAAACTCGGG**TTCGA**GTCCCGGTCTGATGCA

Methanospirillum hungatei strain GP1 chromosome, Reference Sequence: NZ_CP077107.1: 1403216-1403287 tRNA-Gly

GCGTTGATAGTGTAG**TGGT**TATCACTAGGCGT**TGCCA**ACGCCTAAACCCGGG**TTCGA**GTCCCGGTCTAGCGCA

Methanoregula boonei 6A8, complete sequence NCBI Reference Sequence: NC_009712.1: 1259143-1259214 tRNA-Gly

GCGCCGATAGTGT**AGTGGT**TATCACTAGGCGT**TGCCA**ACGCCTAAACTCGGG**TTCGA**GTCCCGATCTGGCGCA

Natrialba magadii ATCC 43099, complete sequence NCBI Reference Sequence: NC_013922.1: c2067059-2067024, c2067011-2066964 tRNA-Gly

GGAAGGGTCGCTCAGC**GGTA**GAGCATCGCGG**TGCCA**CCGCGACGGCAGCACATGCCTCGTGGCG**TTCAA**A
TCCCACCCCTTCCG

Halobacterium salinarum strain 91-R6 chromosome, complete genome GenBank: CP038631.1: c1621908-1621851 tRNA-Gly

CCGACGTGAATAG**AGTGGTA**TCACGTGACCT**TGCCA**TGGTCACAACCTGGG**TTCAA**ATCCCAGCCAGCGCA

Haloferax volcanii strain H98 scaffold_2, whole genome NCBI Reference Sequence: NZ_JAERQU010000003.1: 84353-423 tRNA-Gly

GCGCTGGTAGTGT**AGTGGTA**TCACGTGACCT**TGCCA**TGGTCACAACCTGGG**TTCAA**ATCCCAGCCAGCGCA

Natronomonas pharaonis DSM 2160 complete genome GenBank: CR936257.1: c2368761-2368691 tRNA-Gly

GCGCTGGTAGTGT**AGTGGTA**TCACGTGACCT**TGCCA**TGGTCACAACCCGAG**TTCAA**ATCTCGGCCAGCGCA

Halomicrobium mukohataei strain pws1 Scaffold_1, whole genome shotgun sequence NCBI Reference Sequence: NZ_WOYG01000001.1: c1749471-1749401 tRNA-Gly

GCGGTGGTAGTGT**AGTGGTA**TCACAGGACC**CTGCCA**CGGTCTTAACGGGGG**TTCAA**ATCCCCCCCACCGCA

Haloarcula marismortui ATCC 43049 chromosome, NCBI Reference Sequence: NZ_CP039138.1: 1034685-1034756 tRNA-Gly

GCGCTGGTAGTGTAGT**TGGTA**TCACGTGACCT**TGCCA**TGGTCACAACCTGGG**TTCAA**ATCCCAGCCAGCGCA

Haladaptatus paucihalophilus DX253 B208DRAFT_contig1.1_C, whole genome shotgun sequence NCBI Reference Sequence: NZ_AQXI01000001.1: 1281796-1281866 tRNA-Gly

GCGCTGGTAGTGT**AGTGGTA**TCACGTGACCT**TGCCA**TGGTCACAACCTGGG**TTCAA**ATCCCAGCCAGCGCA

Natrinema pellirubrum DSM 15624, complete genome GenBank: CP003372.1: c854337-854267 tRNA-Gly

GCGCTGGTAGTGT**AGTGGTA**TCACGTGACCT**TGCCA**TGGTCACAACCTGGG**TTCAA**ATCCCAGCCAGCGCA

Natronobacterium gregoryi SP2, complete sequence NCBI Reference Sequence: NC_019792.1: 3540910-3540980 tRNA-Gly

GCGCTGGTAGTGT**AGTGGTA**TCACGTGACCT**TGCCA**TGGTCACAACCTGGG**TTCAA**ATCCCAGCCAGCGCA

Phocaecicola abscessus CCUG 55929, whole genome shotgun sequence NCBI Reference: NZ_HE997184.1: c142655-142583 tRNA-Gly

GCGGAAATAGCTCAGT**TGGTA**GAGCACGACCT**TGCCA**AGGTTCGGGGTCGCGAG**TTTCA**GTCTCGTTTTCCGCT

Bacteroides massiliensis B84634 = Timone 84634 = DSM 17679 = JCM 13223 genomic scaffold aczJI-supercont1.1, whole genome shotgun sequence GenBank: KB905472.1: c1648521-1648446 tRNA-Gly

GCGGAAATAGCTCAGT**TGGTA**GAGCATAACCT**TGCCA**AGGTTAGGGTCGCGAG**TTTCA**GTCTCGTTTTCCGCTCTA

Bacteroides heparinolyticus strain F0111 chromosome, complete genome NCBI Reference Sequence: NZ_CP027234.1: 599826-599898 tRNA-Gly

GCGGAAATAGCTCAGT**TGGTA**GAGCACAACT**TGCCA**AGGTTGGGGTCGCGAG**TTTCA**GTCTCGTTTTCCGCT

Bacteroides fragilis NCTC 9343, complete genome GenBank: CR626927.1: 537974-538046 tRNA-Gly

GCGGAAATAGCTCAGT**TGGTA**GAGCATAACCT**TGCCA**AGGTTAGGGTCGCGAG**TTTCA**GTCTCGTTTTCCGCT

Bacteroides finegoldii strain D53t1_180928_A7 NODE_3_length_206098_cov_23.2893, whole genome shotgun sequence NCBI Reference Sequence: NZ_JADNLJ010000003.1: 45624-45696 tRNA-Gly

GCGGAAATAGCTCAGT**TGGTA**GAGCATAACCT**TGCCA**AGGTTAGGGTCGCGAG**TTTCA**GTCTCGTTTTCCGCT

Bacteroides thetaiotaomicron strain 7330 chromosome, NCBI Reference Sequence: NZ_CP012937.1: 2784339-2784411 tRNA-Gly

GCGGAA**A**TAGCTCAGT**TGGTA**GAGCATAACCT**TGCCA**AGGTTAGGGTCGCGAG**TTTCA**GTCTCGTTTTCCGCT

Porphyromonas asaccharolytica DSM 20707, complete sequence NCBI Reference Seq: NC_015501.1: 369164-369236 tRNA-Gly

GCGGAAGTAGCTCAGT**TGGTA**GAGCACAACT**TGCCA**AGGTTGGGGTCGCGGG**TTTCA**GTCCCGTCTTCCGCT

Table S2. tRNAs from various species; Methanopyrus kandleri other than tRNA-Gly^{GCC}, bacteria, bacteroides and mammals

Methanopyrus kandleri AV19, complete sequence NCBI Reference Sequence: NC_003551.1: c37227-37152 tRNA-Gly^{TCC}
GCGGCCGCAGCTCTAGTCTGGTAGGACGCGGGCCTTCCGAGCCCGTGGCCCCGGGTTCAAATCCCGGCGGGCCGACCA

Methanopyrus kandleri AV19, complete sequence NCBI Reference Sequence: NC_003551.1: 51546-51622 tRNA-Ala^{GGC}
GGGCCCCGAGCTCAGCCCTGGCAGAGCGCGGCCTTGGCGCGCCGTGGCCCCGGGTTCAAATCCCGGCGGGTCCACCA

Methanopyrus kandleri AV19, complete sequence NCBI Reference Sequence: NC_003551.1: 51735-51812 tRNA-Ile^{GAT}
GGGCCCCGAGCTCAGCCAGGTAGAGCGCGGGCTGATAACCCCGAGGTCCCAGGTTCAAATCCCGGCGGGCCCCACCA

Methanopyrus kandleri AV19, complete sequence NCBI Reference Sequence: NC_003551.1: 384999-385073 tRNA-Ile^{CAT}
GGGCCCCGAGCTCAGTCCGGTTAGAGCGCGGGCTCATAACCCCGTGGTCCCAGGTTCAAATCCCGGCGGGCCCCA

Methanopyrus kandleri AV19, complete sequence NCBI Reference Sequence: NC_003551.1: 55068-55107,55184-55221 tRNA-Trp^{CCA}
GGGCCCCGCGGTGTAGCCAGGTCTATCATGCGGGCTCCAAGACCCCGTGACCCGGGTTCAAATCCCGGCGGGCCCCACCA

Methanopyrus kandleri AV19, complete sequence NCBI Reference Sequence: NC_003551.1: c93290-93212 tRNA-Met^{CAT}
AGCGGGGCGGGGTAGCCAGGTCCATCCCAGCGGGCTCATAACCCCGAGGTCCCAGGTTCAAATCCCGGCCCCGCTACCA

Methanopyrus kandleri AV19, complete sequence NCBI Reference Sequence: NC_003551.1: 172302-172377 tRNA-GLN^{TTC}
AGGCGCGTGGGTAGCGGTCTATCCTGCGGGCTTTGGACCCCGCGACCCCGGTTCAAATCCCGGCGGCCCTACCA

Methanopyrus kandleri AV19, complete sequence NCBI Reference Sequence: NC_003551.1: 188239-188327 tRNA-Sec^{TCA}
GGCGCGCGCCACCGGGGTGGTCCCCGGGCGCGGATTTCAATCCCGCGCGCCCCGAGTGGGGCGCGGGGTTCAAATCCCCGCCCA

Methanopyrus kandleri AV19, complete sequence NCBI Reference Sequence: NC_003551.1: c272692-272615 tRNA-Thr^{TGT}
GCCGCGGCAGCTTAGCCCTGGTTAGAGCGCGGGACTTGTAATCCCGTGGTCCCAGGTTCAAATCCCGGCGCGGCTCCA

Methanopyrus kandleri AV19, complete sequence NCBI Reference Sequence: NC_003551.1: 563671-563760 tRNA-Leu^{TAA}
GCGGGGGCGCCCCGAGCCCTGGTCGAAGGGGGCGGACTTAAGATCCGCTGGCCGTAGGGCCGTCCCAGGTTCAAATCCCGGCCA

Methanopyrus kandleri AV19, complete sequence NCBI Reference Sequence: NC_003551.1: c512322-512249 tRNA-Arg^{GCG}
GGCCCCGAGCTCTAGCCCTGGATAGGACGCGGGCTGCGGAGCCCGTAGCCCGGGTTCAAATCCCGGCGGGGCCA

>Bacillus_anthraxis_Ames_chr.trna50:GlyGCC<(746745:746819)<<Gly<(GCC)
GCGGAAGTAGTTCTAGTGGTAGAATAACCTTGCCAAGGTTGGGGTTCGCGGGTTCGAATCCCGTCTTCCGCTCCA

>Bacillus_anthraxis_Ames_chr.trna45:GlyTCC<(543071:543144)<<Gly<(TCC)
GCGGGTGTAGTTTCTAGTGGTAAACAGAGCCTTCCAAGCTCTGGTCGAGAGTTCGATTCTCTTACCCGCTCCA

>Bacillus_cereus_ATCC_10987_chr.trna15:GlyGCC<(150984:151055)<<Gly<(GCC)
GCGGAAGTAGTTCTAGTGGTAGAATAACCTTGCCAAGGTTGGGGTTCGCGGGTTCGAATCCCGTCTTCCGCTCCA

>Bacillus_cereus_ATCC_10987_chr.trna27:GlyTCC<(279452:279522)<<Gly<(TCC)
GCGGGTGTAGTTTCTAGTGGTAAACAGAGCCTTCCAAGCTCTGGTCGAGAGTTCGATTCTCTTACCCGCTCCA

>Bacillus_licheniformis_ATCC_14580_chr.trna55:GlyGCC<(3116926:3116855)<<Gly<(GCC)
GCGGAAGTAGTTCTAGTGGTAGAACACCACTTGCCAAGGTTGGGGTTCGCGGGTTCGAATCCCGTCTTCCGCTCCA

>Bacillus_licheniformis_ATCC_14580_chr.trna23:GlyTCC<(611775:611848)<<Gly<(TCC)
GCGGGTGTAGTTTCTAGTGGTAAACCTCAGCCTTCCAAGCTGATGTCGTGAGTTCGATTCTCATCACCCGCTCCA

>Bacillus_thuringiensis_Al_Hakam_chr.trna84:GlyGCC<(4667815:4667741)<<Gly<(GCC)
GCGGAAGTAGTTCTAGTGGTAGAATAACCTTGCCAAGGTTGGGGTTCGCGGGTTCGAATCCCGTCTTCCGCTCCA

>Bacillus_thuringiensis_Al_Hakam_chr.trna36:GlyTCC<(258057:258127)<<Gly<(TCC)
GCGGGTGTAGTTTCTAGTGGTAAACAGAGCCTTCCAAGCTCTGGTCGAGAGTTCGATTCTCTTACCCGCTCCA

>Bacillus_weihenstephanensis_KBAB4_chr.trna24:GlyGCC<(156592:156663)<<Gly<(GCC)
GCGGAAGTAGTTCTAGTGGTAGAATAACCTTGCCAAGGTTGGGGTTCGCGGGTTCGAATCCCGTCTTCCGCTCCA

>Bacillus_weihenstephanensis_KBAB4_chr.trna37:GlyTCC<(247654:247724)<<Gly<(TCC)
GCGGGTGTAGTTTCTAGTGGTAAACAGAGCCTTCCAAGCTCTGGTCGAGAGTTCGATTCTTACCCGCTCCA

>Bacteroides_fragilis_YCH46_chr.trna52:GlyGCC<(2904501:2904426)<<Gly<(GCC)
GCGGAAATAGCTCAGTGGTAGAGCATAACCTTGCCAAGGTTAGGGTCGCGAGTTCGAGTCTCGTTTTCCGCTCAA

>Bacteroides_fragilis_YCH46_chr.trna55:GlyCCC<(2858492:2858417)<<Gly<(CCC)
GCGGTAGTAGCTCAGTGGCAGAGCGCGGCTTCCAAGCCGAGGTACGAGTTCGACCCCTCGCCTACCGCTCTA

>Bacteroides_vulgatus_ATCC_8482_chr.trna69:GlyGCC<(1160616:1160544)<<Gly<(GCC)
GCGGAAATAGCTCAGTGGTAGAGCATAACCTTGCCAAGGTTAGGGTCGCGAGTTCGAGTCTCGTTTTCCGCT

>Bacteroides_vulgatus_ATCC_8482_chr.trna29:GlyCCC<(3943164:3943236)<<Gly<(CCC)
GCGGCAGTAGCTCAGTGGTAGAGCATCAGCTTCCAAGCTGAGGGTCACGAGTTCGAACCTCGCTTCCGCT

>Bos_taurus_chr3.trna2984:GlyGCC<(86322865:86322938)<<Gly<(GCC)
TCCCTGGTGGTCTAGTGGTTTAAGACTCTGCACCTGCCAATACAGCAGGCCAGGTTCAAATCCCTGGTCAGGGAG

>Bos_taurus_chr3.trna2384:GlyTCC<(68084172:68084244)<<Gly<(TCC)
TCCCTGGTAGTCTAGTGGTTAAGACTCTGTGCTTCCACTGCAGGGGGCTCAGGTTCAAACTCTGGTTGGGGAA

>Homo_sapiens_chr21.trna2:GlyGCC<(18827177:18827107)<<Gly<(GCC)
GCATGGGTGGTTCTAGTGGTAGAATTCTCGCTGCCAAGCGGGAGGCCCGGGTTCGATTCCCGGCCCATGCA

>Homo_sapiens_chr1.trna4:GlyCCC<(17188416:17188486)<<Gly<(CCC)
GCATTGGTGGTTCTAGTGGTAGAATTCTCGCTTCCAAGCGGGAGACCCGGGTTCAAATCCCGGCCAATGCA

Table S3. Percentage of AL-pentamers in rRNAs or mRNAs of proteins participating to the ribosomal translation in *Methanococcus voltae*

***Methanococcus voltae* DNA polymerase gene, complete cds GenBank: L33366.1**

TATGGATTTAGATTATAATTCCAAAGACCTTTGTATTGATATGTATTACAAAAATTGTGGACTAAAAAAA
 CCCGAAATAAATTTACAAAAA**GAATGT**GAGTTTAAACCTTATTTCTATGTAGATACATCAGAACCAAAAG
 AAATTTACGATTATTT**AGATG**GCCTAAAT**CAAGA**AATTGATTTAAAAAAATTAGAACC**TGAAT**TTGAAAA
 TAATACATCCTTAAAGTTTCAGGATTTAATAACAAACATAGAAATTATTGAAAAAATTGTTTATTCTGAT
 TATATTT**TGAATGG**AAAGGATATTTTCAGAAGTAAGTGATTTTAAAAATAAAAAAGAACGAAAAATATGTA
 AAGTCTATGTAAATATCCTAATCACGTGAAAATAATAAGAGAATACTTTAAAGAATTTGGAAAAGTCCT**A**
TGAATTTGACATTCCTTTTTTAAGGAGATATATGATTGAC**CAAGAT**ATTGTACCATCGGCAAAAATATTCC
GAAGATAAT**AAGAT**AGACAACAGTATTCCAGAATTAAATTGTATCGCCTTTGATATGGAACATATTGTA
 AAAAGAACCAAATGCAAAGAAAGACCCTATTAT**AATGGT**TAATTTATTCTCTAAAGACTACCAAAAAAGT
 AATTACCTATAAAAAATTTGAAAATTCGGAATAT**AATGGT**ATGTGTGGATTACGTAA**AAGATGAA**AAAGAA
 TT**GATTCA**AAAAACAATTGAAATTTTGAAACAATATGACGTAATATATACGTAT**AATGG**AGATAATTTTG
 ATTTCCCATATTTAAAGAAAAGAGCAAACATTT**ATGAA**AT**TGAAT**TAGATTTTGACAATGCTTCGAATTC
 TCAACAACCTCAAATCATAAAAAAT**TTCAA**AAGGGGAATCAATAGAAAGAGTAAATTCAGGAATTATC
 CACATTGACCTCTACCCAATAGCTAGAAAATTATTAAATCTTACCAAATATAAATTAGA**GAATG**TAGTAC
 AGGAATTATTTAAATAAAACAAAGAAGCTGTGCGATTATGGCGATATTCCAAAAATGTGGGAAACCG**AAGA**
TACTACATTGCTGAGATATGCCT**ATGAAGATG**CACTATATACCTATAAA**AATGG**GGAATTACTTCTTACCA
 CTTGAAATAATGTTCTCTAGAATAGTTAATCAGCCCTTGTATGATACTAGTA**GAATGAAT**AGTAGCCAA**AA**
TGGTGGAATTTTTATTACTTAAACGGTCTTTTGAACAAAATATGATTTCTCCAAACCGTCC**TTCAAGTTC**
AAGTTACAGAGAACG**TGCCA**AATTCCTCTTACGAAGGTGGTTATGTTAGAGAACCACTAAAAGGT**ATTCAA**
GAAGATATCGTTTCACTTGACTTTATGAGTCTATATCC**TTCAA**TATTGATAA**GCCAT**AATATAAGCCCTG
 AAACGGTCATTTACGAGGAAAAAGAACGAGAAAATATGGAGCTTGGGATAATTCCTAAAACCTTA**ATGA**
ATTATTAAGTCGAAGAAAAACATATTAAATGCTATTAA**AAGATA**AAA**ATTCAA**AAAAATGAGTTT**GATGAA**
 GAATATAGTAGATTAGAACATGAGCAAAAAT**CCATT**AAAGTGCTTGCAAATA**GCCATTATGGT**TATTTAG
CATTCCGATGGCTCGC**TGGTACT**CGGATAAATGTGCGGA**AATGGT**TACAGGACT**TGGTA**GAAAAATAT**AT**
TCAAGAAA**CCATT**GAAAAAGCAGAAGAATTTGGATTTAAGGTAATTTATGCAGATACCG**ATGGT**TTTTAT
 GCAA**AATGGG**ATTATGATAAACT**TTCAA**AAAGGTAAAAAGAAGAAAATGATAAATCTGATAAATTATCTA
 ATTTACCTAAATTATCAAAAGAAGAATTGATAATATTAATAAAAGTTCTTAAAGGAATTA**ATGAAGA**
 ATTACCAGAAGGTATGGAATTAGAATTTGAAGGACATTTTAAAGAGGGTTATTCGTTACAAAAAGAAA
 TATGCGTTAATCG**AAGATGAC**GGGCATATTGTAGTAAAAGGTTTGAAGTTGTAAGACGAGATTGGTCAA
 ACATTGCTA**AAGAT**ACTCAACAAGCCGTTATTAGAGCACTATTGG**AAGATGT**GACGTTAATTTAGCTAA
 AAAAATTATTAATAAATACGATTGATAATTTGAAAAAAGGAAATATTGATAAAATGATTTGCTTATACAT
 ACTCAATTAATAAATAATTTGAGGAATATAAGTCA**ACTGC**ACCACATATTGAAGTAGCTAAAAAATAA
 AGCAGCGTGGTG**ATTCA**GTCCGAGTAGGGGATGTTATAAGTTATATAATTGTCAAAGGTAGCCG**TTCAAT**
 TAGTGAAAGAGCTGAACTATTAGAGTATGCTGGCGATTATGATATTAATTACTACATTGACAATCAGGTT
 TTACCACCCGTTATTAGGATTATGGAGTCTTTGGGAATTTCGG**AAGATGAAT**TGAAAA**ATTCA**GGTAAGC
 AGTTTAAACTCGACCAATTTATGTAA

92 2472 **observed: 3.7% expected: 2.1% ± 0.5 (95% confidence interval)**

***Methanococcus voltae* strain DSM 14649 Ga0466840_02, whole genome shotgun sequence NCBI Reference Sequence:
 NZ_JAGGMO010000002.1: c172863-172297 DNA-directed RNA polymerase subunit D**

ATGCAAAAAGAGGTAAAAAGAACCGGGGAAATTATAACT**ATGAA**AGTTGAAGCTCCTTTGTCTTTTTCCA
 GTGCATTAAGAAGAATAATGATATCAGAATTACCAACTTATGCTATAGAAAATGTTTTTTTCT**ATGAAA**
 CACATCATCAATGTAC**GATGAA**GTTTTAGCTCATAGGTTAGGT**ATGGTAC**CTATAAAAGGAAAACTGTGTA
 AGTTCT**GATGAAT**TAATAACATTTGTAGTATCTAAAGAAGGACCTTGACAGTTTACTCATCAGATTTAA
 AATCAGAAGTTGGGGAACCTGCTTTTGAGAACATACCTTTGGTTAAATTGTCAGAAGGTCAAAAAATTAGA
 AGC**TGAATGT**GAAAGCAT**TGGTAGGA**ACCGGTAAA**ATTCAATTC**AAATGGCAACCATGCAACG**TGGTAT**AT
 ACACAACCTCG**AAGATA**ACCTCGTAGAGTTT**AAGAT**AGAATCACACAAAAATATGGAAGCTGAGGATTTAG
 T**AAGAT**CTGCATTAGAAATTCCTTAAAAATAAGGCAGAAACCTGCTTAAT**TGAAT**TAGAAAGTCATGTATT
 TAATTAA

22 563 **observed: 3.9% expected: 2.1% ± 1 (95% confidence interval)**

Methanococcus voltae A3, complete genome GenBank: CP002057.1: 1754095-1755849 Glycyl-tRNA synthetase

ATGGAAAAATCGATAAAAA**ATG**AAAAATATAATAAAAT**AATGG**ATTTAATCAAAAAGAGAGGTTATTTGT
GGAATTCCTTTGAAATTTATGGGGGAATCGCAGGTTTCTTTGATTATGG**GCCATT**AGGTGCTATTTTAAA
AAATAACATCGTAAATACGTGGAGAAAACATTATATTATTA**ATGAA**GGATTTTACGAAATAGAGGGACCT
ACAATAAATCCTT**ATGAA**GTTTTTAAAAGCTTCTGGACACGTAGACAACCTTTACGGAC**CCATT**AGTACAAT
GTAA**GAATG**TAATGAGTCATTTAGGGCAGACCACATAATTGAAGAAAATGTTGATATTGATACCGAAGG
TAAACACTCG**ATGAAT**TACAAGCTTTAATCACTGAAAACGATATAAAATGTCCACTTTGTGGCGG**TGAA**
TTTAAAGAAGTTGACACT**TTCAAT**CTAATGTTTAA**TTCAAT**TCGGACCAAGCGGTAAAAGAGTAGCAT
TTATGAGACCGGAAACCGCACAAAGGTATATTTATCCAATTTAAAGAATAAACCAAGTTTTTAGAAATAA
ATTACCGTTTCGGGGCTG**TTCAA**ATAGGTAAATCATACAG**GAATG****AA**ATTTAC**CAAGA**CAGGGAGTAATA
AGACTTAGGGAATTTAGCCAAGCTGAAGCT**TGAAT**TTTTTCATACACCCTGATAAAAAAGAACTATGAGAAGT
TTGAACGTGTAA**AAGAT**ATGGTTTT**CCATT**ATTGCC**TTCAA**ATAACCAAGAAGACGAAAACATTGTCTGC
AGAGGATAAAATCGTTAAATGACTTTGGGTGAAGCTGTAGAAAAGGTATCGTAAAAAATAAAGCTATT
GCATACTTTAT**TTCAATA****ACTGCT**AATTTCTTACTCGACGTCGGAAT**AGATG**CAAGTAAAATGAGATTTA
GACAACACTTGCCTAACGA**AATGG**CACACTATGCTTCAGATTGTTGGGATGCAGAACTTTATCCGATAG
ATTTGGATGGGT**TGAATG**TGTGGGTGTAGCAGATAGAACCGATTACGATTTAAATACATATGGCTACA
AGTGGT**G****AAGAT**TTAAGCGTTTTTCGTAGAAT**ATGAA**GAACCTAAGGAAATTGAAACCTTTGATAT**TGAAT**
TAACTTTAAGCACTGGGAAAAACATTTAAAGGAG**ATTCAA**AAATA**ATTCAA**AAATTAATTT**CAGAAAT**
GGATTATCAAAAAATTGAAG**AATGGT**TTCCGGCATCGAAAATA**ATGAA**AAAGGCAATAAATATATTTTA
AAAGCAG**AAGATGGTA**AGGAATTT**TGAAT**TAACAAG**TGAAT**ACCTCACTGCTAAA**AAGAT**AACTAAACAG
TAAACGGCGAAAAAGTAATTCCTCACGTTATTGAGCCTTCATACGGTGTGATAGGATTACTTACTTTGT
TATTGAACACGCT**TTCAATGA**GAAGAAGACAGAACCTTT**ATGAA**ATTAGCACCAAGTGATAGCACCTATT
AAAGCGGGCGTTTT**CCATT**AGTAAATAAAGAGCAAA**TGCCA**ATTATTGCAAGGGGAAATCGAAGAAACAC
TTAGAAATTCGGGAATAATCGCAGAATATGACGATAGCGGTGCTAT**TGGTGA****AAGAT**ATTAA**GAATGGA**
TGAGGT**GGTAC**ACCGTTCTGCATCACCGTTGACGG**TGAAT**CCTTAGATAGCGGAACCGTAACAATCCGT
GAAAG**AGATGAA**AGAACTCAAAA**AAGAT**TAAAAATCGAAGAATTAGG**TGAAT**ATATTAATAATTCCTTAA
AATAA

62 1751 **observed: 5.7% expected: 2.1% ± 0.6% (95% confidence interval)**

Methanococcus voltae strain A 2 Ga0416753_02, whole genome shotgun sequence NCBI Reference Sequence:
NZ_JAGGMU010000002.1: 191261-191872 ATPase subunit E

ATGGGAGCGGAA**AAGAT**AACTTCCAAAATTAAAG**AAGATG**CAAAATTAAAAGCTG**ATTCAA**TCGTAGCTG
ACGCACAAGCTAAATACGAAGCAAC**AATGG**CAGAAGCTAAAGAAGAAGCTGAAAAAAGAAAGCAAGCTAT
TCTTAGAAAAGGAGAAAAAGAAGCAGAA**AATGG**CTGAAAACAGAATTTTAGCTGATG**CAAGAT**TATCCTCA
AAAAAGAAATTGCTTGAAGAAAGAGAAAATACA**ATTCAA**ATGACACTCGAAAAATTAGAGGAAGACTTAA
TGAAATTACCTCAAAA**AAGATGAAT**ACAAAGACATCTTAACAGGT**ATGGTA**ATTAAAGGCGTTCTCTCAAT
CGGTGGCGGAGAATTGACACTCGAA**ATGAACAAGATG**ACTTCGAATTAATCGGTGACGAACTTTGTGG
AAGCTCGAAAAAGAAGTCGAAGAAAAAT**TGAAT**ATTGTAACAGTCTTGAAAAAGGCGAACCAATCGATA
TAATCGGCGGATGTATTGTTAAATCAGC**AGATGGTAC**AAAAGTATCCGACAACAGTCT**TGAATCAA****CATT**
CAAAAGAAATTTAGATAGTGTAAGAGCAAAAATAGCAGATTTATTATTCTAA

31 608 **observed: 5.1% expected: 2.1% ± 1 (95% confidence interval)**

Table S4. rRNAs and rproteins from different species presented in the evolutionary order by G. Caetano-Anollés (given in Harish, A., and G. Caetano-Anollés. 2012. Ribosomal history reveals origins of modern protein synthesis. PLoS ONE. 7:e32776) : Nb of observed pentamers among the pentamer set {ATTCA TTCAA TCAAG CAAGA AGATG GATGA ATGAA TGAAT} (in red), Nb of nucleotides in the sequence (in black), Nb of expected pentamers with its standard deviation (in red), Nb of standard deviations between observed and expected numbers (in blue).

Methanococcus maripaludis strain DSM 2067 chromosome, complete genome GenBank: CP026606.1: 1573556-1574872 S12

ATGGCAATTA AAAACGT**TTCAA**TCGTGTACCCGAATAAATTTGGCGGAGGGATAGCATGTCTTGCAGTGC
ATGTTTTTAACCGCCACTTAAACAA**GTA**CTCAGATATTTCTGCAAAATGCGTATTTTATAGAAAATTACTC
GAAAATTAAGAACAATGACGCTATTTTAATAACTTTACAGTACGAAAACGACTATTTTAACGTTGTAAAA
ATTGTAAAAGAATTA AAAAGCTCAAAATCCAAACGCTATTTTATTGCAGGGGGTCCCTGC**GCCATT****TCAA**
ACCCGCTAACT**ATTCAG**GAATTTTTTGTATGCATTTGTTGTGCGGTGAAATTGAAGGAAGTATACAATGTA
CCAATTGATA**AATGGTA**ATTTTGAATTCCTGGAGTATATGTTCCGAAAAATTATAATA**ATGAA**AAAAAT
AAAAGAATATATCAAAAAAACTCGGGAT**AGATG**ATTATCCAATAGCCCAAGTTACGC**ATGAAT**CGGGAG
CTTATGAAAAGCTTATCTTCTTGAAATTGGAAGGGGTTGCATTAGAACTGTAAGTTTTCATGGCCAA

ATGCATCTACGCTCCTC**CAAGAT**ATCGAAAAATTG**AAGAT**TTAAAGTATCTGGTC**GATGAA**GGATTAAAA
AACACCGATGCAGACAAGGTTTCACTAATTGCCCCATCTGTAAGTGATTATAAGTAT**GTA**CTAGATTTAT
GCGCACACATCTCTGAAAAAATGTGTTAATTTTCGCCTTCATCATTACGGGCAGATACAAT**TACTGATGA**
ACTGCTTGACTTTTTTAAATCTGAAAACATTGACGATAGCACCCGAAGCAGGTAGCGAAGTTCTGCGAAAA
AAAATTGATAAAGGAGTCACTGAAGAAAATATTTTAAATGCAGTTGATATTGCTAAAAACACCGGAATAT
CTACCGTAAACTTTTATTAT**ATGGT**TGGTTTTCT**GATGAAT**CAG**AAGATGA**TATTGAAGAAATAATAAA
TTTAACAAAAAATAAA**AAGAT**AACGTTTCGAAAAGT**AGATG**TTAGCATAAACCC**AATGGT**TCCAAAGCCC
CACACTCCTTTTGAAGACTTT**TGAAT**TTGACATGGACTCAAAAACG**AAGAT**AAAAATATATCGAAAAAAGTC
TGCGAAAAATAAAAGTTAGTGTGATTTTGAAAA**ATTCAA**TTCT**ATGGT**TGCACAGACCGTTCTTG**CAAG**
AGGGGGAAGTGAACT**TTCAA**AAGTTTTAAATGAG**TCAAGA**AACACAATCGAACTTATTAATAATGTGGAT
TTGGATAATTACATTAATAAAATCGAAGAAAAACCTTGGGATTTTATAGAAATCTAA

27 1314 10.26 (3.2) 5.2

Methanococcus maripaludis strain DSM 2067 chromosome, complete genome GenBank: CP026606.1: 387848-388039 S17

TTGGGAAGAATAAGGCAAA**CATTCA**TCAAAAAGAACCGGTGAAGAACTCATTGAGAAATTTGCAGACAAAGT
TCACAAGCGATTTTGAAGAAAACAAAAAAGCTGTTGAAGAAGTTGCAATGATCTCAACAAAA**CCATT**GAG
AAACAGAATAGCAGGTTATGT**TACTGCT**TAAAGTAAAAAAA**ATGAA**CGCATAA

2 188 1.47 (1.2) 0.44

Methanococcus maripaludis strain OS7mut1 chromosome, complete genome NCBI Reference Sequence: NZ_CP020120.1: 1380141-1380545 S9

GTGAAAGTAGTCCACACAGTAGGTAAAAGAAGAACCGCAATTG**CAAGA**GC**TACTGC**AAAAGAAGGAAGCG
GTAAATCAGAATTAACAAAAAA**CCATT**AGAATT**AATGGA**ACCTAAATACATCAAA**ATGAA**ATT**AATGGA**
ACCTGTTATCTTAGCAGGCGAAGCTTTAAGCAACATTGACGTAGATATTGATGTTAAAGGTGGCGGTATA
GTAAGCCA**AATGGAT**GCTA**CAAGA**ACCGCTTTAGGTAAAGCTATTGT**TGAAT****TACTG****GTAAATGG**ATT
TGAAAGAAAAATTCTTAAGTTACGACAGAACATTACTTGTCTAGTGACG**CAAGA**AGAACCGAACCACACAA
ACCAAGTAAGTCCTCAAAAGGTC**CAAGA**GCATAAAGACAGAAATCATACAGATAA

6 401 3.13 (1.77) 1.62

Methanococcus maripaludis strain DSM 2067 chromosome, complete genome GenBank: CP026606.1: c1304042-1303581 L22

ATGGCTAAACTAAAATACAAAGTTGAAGCTGACCCAAAGAAC**ACTGCAAG**ACTATGGGCAGAACTTTAA
GAATA**TCAAGA**AAACATGCTATTGAACCTTTCAGAGAATTAAGTG**GAATG**AAATT**AGATG**CTGCTGTTGC
ATACTTAAACAGAGTAATCGCATTAGAACTCCTGTT**CCATTCAA**AGTTCACAACAAAGACTTACCACAC
AGAAAAGGAAAAATTGGAACACACTCAGGTAGATTCCACAGAAAAGCTTCGTTAGAAATCTTAAACGTTT
TAGATAACGCTAAGAAAAACGCTGAGCAAAAAGGATTAAACACAGAAAAATTAAGAATTAAACACATATC
ATCAAACAGAGGCTTCACAATTAA**AAGAT**ATAT**TGCCAAG**GCATTTGGTAGAGCAAGCCCTAAAAAC**CAA**
GAAACAATACATATCCAAGTAATCCTTGAAGAATTCTACTAA

7 1.5% 458 3.6 (1.9) 1.8

Methanococcus maripaludis strain OS7mut1 chromosome, complete genome NCBI Reference Sequence: NZ_CP020120.1: 1575149-1576153 L3

ATGGGT**ATGAA**AAAAAATAGACCTCGTAGAGGTTCCCTAGCATTTAGCC**CAAGA**AAAAAGAGCTAAGAAAT
TAGTTCCTAAATCAGATCATGGCCTGCAGATAAAAAAGTAGGGCTCCAAG**CATT**CCTGTATACAAAGC
AGGAACAACACATGCTTTATTGATCGAAAACAACCTAAAAGTCCAACAGCAGTCAGGAAGT**ATTCA**CA
CCCGTAACTGTATTAGAAACACCGGACGTTACCGTAGCAGGAATCAGACTCTACGAAAAACAACCAAAG
GGTTAAAGCATTAACTGAAGTATGGGCTGAACAGTTAGATAGCGACTTAGGAAGAAAAATTAACCTTTAGT
TAAGAAAGAAGAGAAAAAACTGTTGACGCTTTAGACGCAGTTTTAGAAAAAGCTACAGAAGTAAGAGCT
ATTGTTCCACACAACCCCAAAACTACAGGAATTCCAAGAAAAAACCTGAAGTTGTTGAATTAGAATCG
GTGGATCATCAGTTGCTGA**AAGAT**TAGCTTATGCTAAAGAAATCTTAGGCAAAACACTCGCAATTAGCGA
TGTTTTTCGAAGCTGGCGAAATTATCGATACCTTAGCAATTACAAAAGGTAAAGGATTCCAAGGATCAGTT
AA**AAGATG**GGGAATCAAAG**TTCAA**TTTGGAAAAACACCAAGAAAAAGGTGT**TGGTA**GACACACAGG**TTCAA**
TCGGCCCATGGAGACCTAGAAGAGTTATGTGGACTGTA**CCATT**ACCTGGTCA**AATGG**GTTTCCACCAAAG
AAC**TGAAT**ACAACAAAAGAATCTTAAATTAGGCAGCGAAGGCGCTGAAATTACACCTAAAGGCGGATTC

TTAAACTACGGTGTCTGTTAAAAACGGCTACGTTGTAGTTAAGGGAAGCTG**TTCAAG**GTCTGCAAA**AAGAT**
TAGTAGTTTTTAAGAGGATCTGTAAGAGCTGCTGAAGACAAATTCGGCTTACCTGAAATTGCTTACATCAG
TACAGAATCCAAACAAGGAACTAA

9 1001 7.8 (2.8) 0.4

Methanococcus maripaludis L4

ATGAATGTTAAAGTTTACAATTT**AGATG**CTCTGAAAAGGGAGATAT**TGAAT**TACCCGCTGTATTTGAAGCAGAA
TACAGGCCTGATTTAATTAAAAGGGCAGTAATCTCATCATTAAACCGCAAAATTACAACCAAAAGGTTGCGATGCA
TTTGCAGGATACAGAACTTCAGCAAAATCAATCGGAAAAGGACACGGTAAAGCTAGAGTTAGAAGAACCGCACAA
GGTGCTGGTG**CATTG**CTACCCCAAGCAGTTGGTGGAAGAAGAGCTCACCTCCTAAAGTTGAAAAAATCTTATTC
GAAAGAATAAACAGAAAAGAAAAATTTAAAGCATTAGCAAGTGCTATTGCAGCTTCAGCAAAACCTGAAATCGTT
TCAGCAAGAGGCCACAAAATTGAAGGTGTTCTTCATTACCTTTGGTTGTTAACGCAGAAATTTGAGAGCCTTGTT
AAAACAAAAGAAGTTTTAGAAGTT**TTCAA**AACTTTAAAATT**AGATG**CTGACTTAGAAAAGAGCTAAAAGACGGCGTT
AAAATTAGAGCAGGAAGGGGAAAATTAAGAGGAAGAAAATACATAAAACCAAAAAGTGTTTTGGTTGTTGTAGGC
GATGTATGTGAAGCAAT**TACTGCATCAAGA**AACCTCGCAGGTGT**AGATG**TAATTACAGCTAACGATTTAAGCGCA
ATACACATTGCACCAGGAAC**AATGG**CTGGA**AAGAT**TAACACTCTGGACTGAAAA**TGCCATT**GAAAAAATAAACGGA
AGATTTTTAA

9 755 5.9 (2.4) 1.3

Methanococcus maripaludis L1

ATGGACAGTGAAAAAA**TACTG**AACGCAGTGAAGGAGGCTCGAACTCTTGCTAAGCCGCGAAACTTCACGCAATCC
GTTGATCTCATCGTCAACTTAAAGAGTTGGATCT**TTCAAGA**CCTGAAAACAGGTTGAAAAGAGCAAAATCGTTTTA
CCAAGTGGAAGAGGTAAAGACGTAGCGATTGCGGTTATTGCGAAAGGTGACTTGGCTGCTCAAGCAG**AAGAT**ATG
GGCCTCACTGTAATAAGACAGGAAGAATTAGAAGAATTAGGTAAAAACAAAAAACAGCTAAAAAAATTGCTAAC
GCACACGGCTTCTTTATCGCT**TCAG**CTGATATGA**TGCCATT**GGTCGGTAAATCATTAGGTCCGGTTCTAGGTCTT
AGAGGTAAAA**TGCCA**CAACCAAGTGCCTGCAAATGCAAACTTAGCTCCATTAGTTGCTAGATTCCAAAAAACCGTT
GCAATAAACACCAGGGATAAAGCATTATTCCAGGTATACATCGGAACAGAATCAATGAGCGAC**GATGAA**CTTGCT
GCAAATGCCGAAGCAATTTTAAACGTAGTGTCTAAAAAATACGAAAAAGGACTCTACCAGGTTAAAAATGCATTT
ACAAACTTAC**AATGG**GCGCAGCTGCTCCTATTGAGAAATAG

6 638 5 (2.23) 0.45

Methanococcus maripaludis strain OS7mut1 chromosome, complete genome NCBI Reference Sequence: NZ_CP020120.1: 1464911-1465375 L30

ATGGCTTACGCAGTTGTAAGAGTCAGAGGAAGCGTAGGTGTTAGAGGAGACATCGCAGATACC**ATGAAGA**
TGTTAAGACTACACCGTGTAACCACTGTGTTATAATTCCAGATACAGAACACTACACTG**GAATG**ATAAA
AAAAGTTAAAGACTATGTAACCTACGGTGAAATCGATA**AAGAT**TACACTTGTAGCTTTAATTTTAAAAAGA
GGAAGACTTCTTGGAACAA**AAGAT**TATCTGAAGAACTCGTAAAAGAGTTAACAGAACTTCCAGTTGAAG
AGTTAGCTGAAAAAGTAATTGCTGGCGAAATTTAAATAAAAAGACACCAATAAAACCTGT**ATTCA**GATT
ACATCCTC**CAAGA**AAAGGATATGATAGAGCAGGAGTTAAAAAAGGATTCTCAATTGGTGGTGCTTAGGT
TACAGATC**TGGT**AAAATCAACGATTTATTGAACAAAATGATGTAA

4 461 3.6 (1.9) 0.2

Homo sapiens ribosomal protein S12 (RPS12), mRNA NCBI Reference Sequence: NM_001016.4

CTCTTTCC**CTGCC**GCCGCCGAGTCGCGCGGAGGCGGAGGCTTGGGTGCG**TTCAAGAT****TCAA**CTTCACCCG
TAACCCACC**GCCAT**GGCCGAGGAAGGCATTGCTGCTGGAGGTGT**AATGG**ACGTTAA**TACTG**CTT**CAAG**
AGTTCTGAAG**ACTGCC**CTCATCCACGATGGCCTAGCACGTGGAATTCGCGAAG**CTGCCA**AAGCCTTAGA
CAAGCGCCAAGCCATCTTTGTGTGCTTGCATCCAACCTGT**GATGA**GCCTATGTATG**TCAAG**TTGGTGAG
GCCCTTTGTGCTGAACACCAATCAACCTAATTAAGGTT**GATGA**CAA**CAAGA**AACTAGGA**GAATGG**GTAG
GCCTTTGTAAATTGACAGAGAGGGGAAACCCGTAAAGTGTTGGTTGCAGTTGTGTAGTAGTTAAGGA
CTATGGCAAGGAGTCTCAGGCCAAGGATGTCATTGAAGAGTAT**TTCAA**ATG**CAAGA****ATGAA**GAAATAAA
TCTTTGGCTCACA

13 499 3.9 (2) 4.6

Homo sapiens ribosomal protein S17 (RPS17), transcript variant 1, mRNA NCBI Reference Sequence: NM_001021.6

GTTTCCTCTTTTACCAAGGACCCGCCAACATGGGCGCGTTCGCACCAAAACCGTGAAGAAGGCGGCCCC
GGTCATCATAGAAAA**GTACT**TACACGCGCCTGGGCAACGACTTCCACACGAACAAGCGCGTGTGCGAGGAG
ATC**GCCATT**ATCCCCAGCAAAAAGCTCCGCAA**CAAGAT**AGCAGGTTATGTACGCATCT**GATGAA**GCGAA
TTCAGAGAGGCCCAGTAAGAGGTATCTCCA**TCAAG**CTGCAGGAGGAGGAGAGAGAAAAGGAGAGACAATTA
TGTTCTGAGGTCTCAGCCTTGATCAGGAGATTATTGAAGTAGATCCTGACACTAAGGAAATGCTGAAG
CTTTTGACTTCGGCAGTCTGTCCAACCTTCAGGTCAGCCTACAGTTGG**GATGAATTTCAA**AACGC
CTCGGGACCTGTT**TGAAT**TTTTTCTGTAGTGCTGTATTATT**TTCAA**TAAATCTGGGACAACAGCCTT

8 1.7% 484 3.8 (1.9) 2.2

Homo sapiens ribosomal protein S9 (RPS9), transcript variant 1, mRNA NCBI Reference Sequence: NM_001013.4

CTCTTTCTCAGTGACCGGGTGGTTTGCTTAGGCGCAGACGGGGAAGCGGAGCCAACA**TGCCA**GTGGCCCC
GAGCTGGGTTTGTGCGAAACTTATGTGACCCCGCGGAGACCCTTCGAGAAATCTCGTCTCGAC**CAAGAG**
CTGAAGCTGATCGGCGAGTATGGGCTCCGGAACAAACGTGAGGTCTGGAGGGTCAAATTTACCCTGGCCA
AGATCCGCAAGGCCGCCCGGGA**ACTGCT**GACGCTT**GATGA**GAAGGACCCACGGCGTCTGTT**CGAAGGCAA**
CGCCCTGCTGCGGCGGCTGGTCCGCATTGGGGTGCTG**GATGAGGGCAAGATGAA**GCTGGATTACATCCTG
GGCCTG**AAGAT**AGAGGATTTCTTAGAGAGACGCCTGCAGACCCAGGTCT**TTCAAG**CTGGGCTTGCCAAAGT
CCATCCACCACGCTCGCGTGCTGATCCGCCAGC**GCCAT**ATCAGGGTCCGCAAGCAGGTGGTGAACATCCC
GTCCTTCATTGTCCGCCTGGATTCCCAGAAGCACATCGACTTCTCTCTGCGCTCTCCCTACGGGGGTGGC
CGCCCGGGCCGCGTGAAGAGGAA**GAATG**CCAAGAAGGGCCAGGGTGGGGCTGGGGCTGGAGACGACGAGG
AGGAGGATTAAGTCCACCTGTCCCTCCTGGGCTGCTGGATTGTCTCGTTTTCT**CTGCCA**ATAAACAGGAT
CAGCGCTTTACAA

9 709 5.5 (2.35) 1.49

Homo sapiens L3 mRNA, complete cds GenBank: AY598322.1

ATGATGGCCACTCCGAACCAGACCGCCTGTAATGCAGAGTCAACAGTGGCCCTGGAGGAGGCCAAGACCT
CTGGTGCCCCGGGGAGCCCCCAAACACCCCTGAGCGTCATGACTCTGGTGGTTCC**CTGCC**CCTGACACC
GCGGATGGAGAGCCACTCAGAG**GATGAAGAT**CTTGCTGGGGCTGTGCGTGGCCTGGGCTGGAACAGTAGG
AGTCCCCGGACCCAGAGCCCAGGGGGCTGCTCAGCGGAGGCTGTGCTGGCCCCGAAGAAACACCGTCGGC
GCCATCGAAGCGCAAAAGGCACTGGCGACCCTACCTGGAGCTGAGCTGGGCTGAGAAACAACAGCGG**GA**
TGAGAGGCGAGCCAGAGGGCCTCCCGGGTCCGCGAAG**AGATG**TTCCCAAAGGCCAGCCCGTGGCCCCC
TACAACACCACCCAGTTCTCT**GATGAATG**ACAGGGACCCGGAGGAGCCCAACTTGATGTGCCCATGGGA
TCTCCCACCCAGGTTCCAGTGGGGAGAGTGAGGCCGGGGACAGTGATGGGCGGGGCCGAGCGCACGGTGA
GTTCCAGCGGAAGGACTTCTCTGAGACTTACGAACGCTTCCACACCCGAGAGCCTGCAGGGCCGACGCAAG
CAGGAGCTGGTGCAGACTACCTGGAGCTGGAGAAGCGGCTGTGCGAGGCGGAGGAGGAGACTAGGAGGC
TGCAGCAGCTGCAGGCGTGACCGGCCAGCAGTCCTGCCGCCAGGTGGAGGAGCTGG**CTGCC**GAGGTCCA
GAGGCTCCGGACCGAAAACAGCGGCTTCGTACAGGAGAACC**AGATG**TGGAACCGAGAGGG**CTGCC**GCTGT
GATGAGGAGCCGGGTACCTAG

9 1.3% 657 5.1 (2.3) 1.7

Homo sapiens ribosomal protein L22 (RPL22), mRNA NCBI Reference Sequence: NM_000983.4

CCTTTCTAACTCCGCTGCC**GCCAT**GGCTCCTGTGAAAAAGCTTGTGGTGAAGGGGGGCAAAAAAAGAAG
CAAGTTCTGAAGTTCACTCTTGATTGCACCCACCCTGTAG**AAGATG**GAATCATGGATG**CTGCCA**ATTTTG
AGCAGTTTTTG**CAAGAA**AGGATCAAAGTGAACGGAAGAGCTGGGAACCTTGGTGGAGGGGTGGTGACCAT
CGAAAGGAG**CAAGAGCAAGAT**CACCGTGACATCCGAGGTGCCTTTCTCCAAAAGGTATTTGAAATATCTC
ACCAAAAAATATTTGAAGAAGAATAATCTACGTGACTGGTTGCGCGTAGTTGCTAACAGCAAAGAGAGTT
ACGAATTACGTTACTTCCAGATTAACCAGGACGAAGAAGAGGAGGAAGACGAGGATTAATTTTCAATTTAT
CTGGAAAATTTTGATGAGTTCTTGAATAAAACTTGGGAACCAA**AATGGT**GGTTTATCCTTGTATCTCTG
CAGTGTGGATTGAACAGAAAATTGGAAATCATAGTCAAAGGGCTTCCCTTGGTTCGCCACTCATTTATTT
GTAACTTGACTTCTTTTTTTTTCTGCTTAAAAAT**TTCAA**TTCTCG**TGGTA**ATACCAGAGTAGAAGGAGAG
GGTGACTTTACCGAACTGAC**GCCATT**GGGGAGGC**AGATG**CGGGTGTGGAGGTGTGGGCTGAAGGTAGTG
ACTGTTTGATTTTAAAAAGTGTGACTGTGAGTTGTATCTGTTGCTTTTCTCAATGATTCAGGGATACAAA
TGGGCTTCTCT**CATTCA**TAAAAAGAAAACGCGACATCTTTCT**AAGAT**TCTCTGTGGGAAAATGACTGTCA
ATAAAATGCGGGTTTCTGG**GCCATT**CGTCTTACTTTTCAATTTTGTATTACAAATTTCTCTTGACGCACAC
AATTATGTCTGCTAATCCTCTTCTTCCTAGAGAGAGAACTGTGCTCCTTCAGTGTTG**CTGCCAT**AAAGG

GGTTTGGGGAATCGATTGTAAAAGTCCCAGGTTCTAAATTAACATAATGTGTACAGAA**ATGAA**CGTGTAA
 GTAATGTTTCTACAGGTCTTTGCAACAACTGTCACTTTTCGTCTCCAGCAGAGGGAGCTGTAGGAATAGT
 GCTTCC**AGATG**TGGTCTCCCGTGTGGGGCCAGC**AATGG**GGGGCCCTGATGC**CAAGA**GCTCTGGAGGTTC
 TTGAAAGAGGGGACACGAAGGAGGAGTGAAGTGGGAAGCCTCCCA**TGCCA**AGGAGGTGGGAGGTGCCCTGG
 AAATAGCTGCCTCA**TGCCA**CTTAG**GCCAT**GAAGTGGATTTAATGTCAAGTGGT**TGCCA**CAGTGCAGAGGCT
 AGACAACTGAAAGGGGCTACCAAGGCTGGGAAAAAATGCAATTGTTGCTGTGAGTGAAGTGTGAAAGACT
 CTGGTGCCTTGTGGTGCCTTCTGAA**ATTCAA**ACAGTAATGCAAAAGTGTCTGCATTAGAATTTACGGTG
 TCTAAA**ATTCA**TGTTTTTAAAGAGCTTGCCTAC**AGATG**GTTTCCACACTTGAAATGTGCCCTGCGAGT
 TGCATAGCTGGAAG**TTCAAT**GTCTCAGTCTACCTTGGCTCCCATTAACATTTGGTGTCTGTGGATTGA
 GTTGAACGTGTTGAGGCTTTGCAATTTCACTTGTGTTAAAGGCTCTGGCATTTTT**CCATT**TCATGCAAA
 TTTCTTTGAAGCAGAATTGCTTGCATATTTCTTCT**CTGCC**GTACAGAAAGCAGAGTTTCT**TTCAA**ACTT
 CACTGAGGCATCAGTTGCTCTTTGGCAATGTCCCTTAACCATGATTATTAACATAAGTTTGTGGCTTGAGT
 TTACAAATTCTACTTGTGTCATTGATGTTCCCATGTAGTAAGTCATTTTTAGTTTGGTTGTGAAAAAACC
 CTGGGCTGAAGTTGGCATTTCAGTTAAAAGAAAAAAGAACTAGTCCAGATTTGAAAACTGTAAATAA
 AATTGAACTCACTGGTTTTCTATGTCTTT**TTGAA**CTCTTGTAAATCGAGTTTTGATCATATTTCTATTA
 AAGTGGCTAACACCTGGCTACTCT**TACTG**TA

16 0.7% 2057 16 (4) 0

Homo sapiens L4 mRNA, complete cds GenBank: AY598323.1

ATGTTTGGTTTTTCACAAGCCA**AAGATG**TACCGAAGTATAGAGGGCTGCTGTATTTGCAGAGCTAAGTCCT
 CCAGTTCTCG**ATTCA**CTGACAGTAAACGCT**ATGAA**AAGGACTTCCAGAGCTGTTTTGGATTGCATGAGAC
 TCGTTTCAGGAGACATCTGCAATGCCTGTGTCTCTGCTTGTGAA**AAGATG**GAAGAAGTTGCCAGCAGGATCA
 AAAAAAACTGGAATCATG**TGGTAGATG**CAAGGGCTGGACCCAGTCTAAAGACTACATTGAAACCAAAGA
 AAGTGAAGTCTATCTGGGAACAGGATAAAAAGCAACCAGATCAGTAA**ACTGC**AGAAGGAATTTAAACG
 TCATAATTCTGATGCTCACAGTACCACCT**CAAGT**GCCTCCCCAGCTCAATCTCCTTGTACAGTAACCAG
 TC**AGATG**ACGGCTCAGATACAGAGATGGCTTCTGGTTCTAACAGAACACCAGTTTTTCTTTTATGATC
 TCACTTACTGGAAAAGACAGA**AAGAT**ATGTTGTGGGATCATCTATAAAGGCCGTTTTGGGGAAGTCCTCAT
 TGACACACATCTCT**TTCAAG**CCTTGCTGCAGCAATAAGAAAGCAGCTGCTGAGAAGCCAGAGGAGCAGGGG
 CCAGAGCCT**CTGCC**CATCTCCACTCAGGAGTGGTGA

10 662 5.17 (2.3) 2.1

PREDICTED: Homo sapiens mitochondrial ribosomal protein L1 (MRPL1), transcript variant X2, mRNA NCBI Reference Sequence: XM_047416090.1

GCAATCCGGAGTGCCCAACATGGCGGCGGCCGTAAGGTGCATGGGTAGAGCCTTGATACATCATCAAAGG
 CATAGCCTTTT**CAAGATG**TTTATCAGACATCACTTTGTTCTTGTCTGTAAACATCCGAGTGCCCAACA
 GACATTTTGTCTGCTGTACAAAGTCTGCAAAGAAAAACAAAAAAGGTGCTAAAGAAAAAACACC**AGATGA**
 GAAAA**AAGATGAA**ATAGAAAAAATAAAAGCATATCCCTATATGGAAGGCCGAACCTGAG**GATGA**TGTCTAT
 TTAAACGCTTATACCCGAGACAGATATATGAGGTGGAGAAAGCTGTTCACTTACTTAAGAAAT**TTCAA**A
 TTCTTGACTTTACTAGTCCAAAGCAAAGTGTATCTTGATTGACACTGGATATGGCACTGGGAAAGAA
 GAAAAACGTGGAG**GCCATT**TACCAGTGTTCCTTAGTT**TGCCAT**AC**CCATT**TGCTTCCGAAATCAATAAAGTT
 GCTGTATTTACAGA**GAATG**CATCAGAGGTCAAAATAGCGGAAGAA**AATGG**AGCTGCATTTGCAGGAGGCA
 CTAGTCTGATACAGA**AAGAT**TTGGGAT**GATGAA**ATTGTTGCAGACTTTTACGTAGCTGTTCCAGAAATAAT
 GCCTGAAGTAAATCGATTAAGGAAGAACT**TGAAT**AAAAAATATCCAAAGCTTTCTCGAAATTTTAA**ACTG**
CCTATGGGTAGTTGCTTTCACTTATAGCTGTAGTGCTCCTGAGCCAACTGAACAACTGCTGAGATAATCC
 TAGGAATAATGTATGCTGAGGGTATGCCGTGCTCTCTATGTAATGAGTTCCTCACATTG**ATTCA**TCTATG
 GAAGACTTGCTTGAGC**ATGGT**GGGAGTGATGTCTTTCAGTGGCACAGTCTT**CATT**CATTTCTTACCCT
 TTAAATAATTACATTTTTTCACTGCATGGAATTAAATTTT

14 946 7.4 (2.7) 2.4

Homo sapiens ribosomal protein L30 (RPL30), mRNA NCBI Reference Sequence: NM_000989.4

CCTTTCTCGTTCCCCG**GCCAT**CTTAGCGGCTGCTGTTGGTTGGGGGCCGTCCCGCTCCTAAGGCAGG**AAG**
ATGGTGGCCGCAAGAAGACGAAAAAGTCGCTGGAGTGCATCAACTCTAGGCTCCAACCTCGTT**ATGAAA**A
 GTGGGAAGTACGTCCTGG**GGTACA**AGCAGACTCTG**AAGATGA**TCAGACAAGGCAGGAAAGCGAAATTTGGTCAT
 TCTCGCTAACAA**ACTGCC**CAGCTTTGAGGAAATCTGAAATAGAG**GTACT**ATGCTATGTTGGCTAAAAGTGGT
 GTCCATCACTACAGTGGCAATAATATTGAACTGGGCACAGCATGCGGAAAATACTACAGAGTGTGCACAC

TGGCTATCATTGATCCAGGTGACTCTGACATCATTAGAAGCA**TGCCA**GAACAGACTGGTGAAAAAGTAAAC
CTTTTCACCTACAAAATTTACCTGCAAACCTTAAACCTGCAAAATTTTCCTTTAATAAAAATTTGCTTGT
TTTAAAAA

2 494 1.9 (1.4) 0.7

Table S5. Glycyl-tRNA ligase sequences from different species with after the sequence: Nb of observed pentamers among the pentamer set {ATTCA TTCAA TCAAG CAAGA AGATG GATGA ATGAA TGAAT} with percentage among the set of pentamers of the sequence (in red), Nb of nucleotides in the sequence (in black), Nb of expected pentamers with its standard deviation (in red), Nb of standard deviations between observed (in blue) and probability of this event (in green).

Acinetobacter baumannii strain ATCC BAA-1790 chromosome, complete genome GenBank: CP042841.1: c395610-393541 Glycyl-tRNA synthetase subunit beta

ATGAGCAAACA**TACTG**TTTTATTTCGAACCTTGGCTGTGAAGAGCTTCCACCAAAAAGCTTAAAAACATTAC
GTGATGCACTACAAGCAGAACTGTAAAAGGCTTAA**ATGAAG**CAGGTTTAAACTTCGCATCTGTTGAAGC
TTATGCCGCTCCACGTCGTTTAGCGCTTAAATTTGT**AGATGTAGATG**CTGC**TCAAG**CGGATACACAAAAA
CGTTTTGACGGCCCTGCTGT**TTCAAG**CAGCTTACGATGCAGAAGGCAAAACCGACTAAAGCACTTGAAGGTT
TTATGCGCGG**TCAAG**GTAT**TACTG**TTGATCAGCTTTCTA**CATTC**CAAGCAGGTAAAGTTGAAAAAGTTTG
CTATTTAA**AAGATG**TGAAAGGCCAAAGCCTTGACACTTTGCTTCCACAGATTTTACAA**ACTGC**GTTAGAC
AATTTACCAATTGCTAAGCGCATGCGTTCTGCTGCAAGCCGTACCGAATTTGTACGTCCAGTGA**AATGGG**
TTGTCTTACTCA**AAGATG**ACCAAGTTATTGAGGCGACT**ATTCAAGAT**CACAAAGC**TGGTA**ATGTGACTTA
TGGTCACCGTTTTCCATGCACCTGAAGCTGTGACCTTAGCGCATGCAATGACTATTTAGCTGCGTTAGAA
AAAGCATATGTTGTAGTGAACCTTTGAAAAGCGTCAAACAAC**ATTCAAGA**ACAAGTAAAAAGTTAGCT**G**
ATGAAGTGAACGCA**ACTGC**GATTGTTCCAGCTGACTTACT**GATGAA**GTAACAAGTCTGGT**TGAATGG**CC
TGTAGCTTTACGTGCAACTTTTGAAGAACGCTACTTAGCTGTTCC**TCAAGA**AGCGCTCATTACGACAATG
CAAGACAACCAGAAGTATTTCTGTTTGATCAATGCGGAAGGTAAATTACAGCCTTACTTCAT**TACTGTTT**
CAATATTGAGTCTAAAGACCCGACTCAAATTATTGAAGGTAAACGAAAAAGTTGTTCGCCCTCGTTTGTC
TGATGCTGAGTTCTTCTTCTTA**CAAGAT**CAAAAGCAACCGCTTGCACTCTCGTAAAGAAAAATTAGCAAAC
ATGGTATTCCAAGCACAGCT**TGGTAC**GCTTTGGGACAAATCAACACGTATTGCAAACTTGCAGTTGCAT
TATCT**TTCAAT**CACTGGTGCAAAACCCAGCGGATGCTGAAAAAGCTGCATTACTTGCTAAATGTGACTTAAC
TTCTGAACCTTGTAGGTGAGTTCCCCGAAC**TTCAAG**GTATTGCG**GGTACT**TACTACGCGCGTATTGAAGGC
GAAAACACTGAAGTTTCAGAAGCTTTAGGTGAACAATACTTACCAAAATTTGCGGGTGATGTTTACCTA
AAACCAAAACA**GGTACAA****CCATT**GCCCTTGCTGACCGTTTAGATACGCTTGT**TGGTA**TTTTCGGTATTGG
ACAAGCGCCTACAGGTTCTAAAGAC**CCATT**TGCTCTTCGCCGTTTCAGCAAT**TGGTA**TCTTACGCTTAATC
ATTGAAA**ATGAAT****AGATG**TAACAATTGAAGAGCTAGTTAATCTTGCACT**TTCAAG**GTTATGGCGATATCG
TTAAAGACCACGACAAAACCTCGTGCTGATGCGGTTG**CATTC**TTGGAAGGTCGTTACCGTGCTAAGT**ATGA**
AGACCAAGGCGTTGCTGTTGACGTCC**TTCAAG**CAG**TTCAAG**CTCTTGCACCAAAATCTCCACTTGATTTT
GACAAGCGCGTAAATGCTGTTAACCACCTCCGTACATTACCTGAAGCTGCTGCACTTGCTGCTGCAACA
AACGTGTTGCTAA**CATT**TTGCAAAAGAAACAGACCAAGGTTCTGTGCTTGAAGCGAACTTGTT**G**
AGATGCTGAAAAAGCATTATTGCTGAGCT**TTCAAG**CAGTGACACCTGTGGTTGA**CCATT**TTGTCAGCA
AAAGACTACACAG**CTGCC**CTTTCTAAGCTTGCAGCACTTCGCGCGCGATTGATG**CATT**CTTTG**TGGT**G
TAATGGTAATGGCAGATGATGCTGACCTAAAGCAAACCGTTTACGTTTATTAGCTCAGTTACGTAACTT
GTT**TACTGC**TGTGGCTGATGTGAGTGTGTTGCAGGGTTAA

44 2.1% 2066 16.1 (4) 7 3 10⁻¹³

Striga asiatica UVA1 DNA, scaffold5, whole genome shotgun sequence GenBank: BKCP01008292.1 glycyl-tRNA synthetase

ATGAATGTCGC**TCAAGA**AGCCCAGAACCAGGGAGG**CATTC**CGGCACGCCGTCGTCAACACCCTCGAACGCC
GTTTGTTTTACATACCTTCC**TTCAAGAT**CTACCGCGGTGTCGCCGGCCTTTATGATTATGGGCTCCGGG
CTGCGCCGTTAAGTCTAACGTCCTCGCCTTCTGGCGTCAGCATTTTGTCTGGAGGAGAACATGCTAGAA
GTTGATTGCCCATGTGTCACTCCAGAGGTTGTGCT**TCAAG**GCATCAGGCCACGTGGACAAGTTCACCGACC
TT**ATGGT**CAAAGACGAAAAGACTGGAACATGTTACCGAGCAGATCATCTGC**TCAAG**GATCATTGCAAGA
AAAGCTCGAGAAGGACCCTAATCTCAGTGCTGATAAGGCAACTGAACCTTAGGCATGTGCTTGCTGTGCTG
GATGATCTTTCTTCAAGGAGCTTGGTGCTAAAATAAAGAGTTATGGAAT**TACTGCT**CCGGACACAAAGA
ACAGTCTGTGACACCCTTACCCTTTTAACTAATGTTTCAGACATCAATTGG**GCCAT**CTGGCTTGAGCAG
TGGGTATATGCGTCCTGAGACGGCTCAGGGTATTTTTG**TGAAT****TCAAG**GACTT**GTA**CTACTAT**AATGGC**
AACAAGCTTCCTTTTGCAGCTGCACAGATTGG**TCAAG**CTTTCCGAAATGAGGTCTCGTATTCTACCATGA
TGCTTGCAATTGCTACTTGCTTTCTTTGAGTATCTTCTAGTGTAGCCTGGGACCTGGAG**AAGAT**CTCTCC
TAGGCAAGGTCTACTTAGGGTCCG**TGAAT****TCA**CACTGGCAGAGATTGAGCACTTTGTGGATCCTGAGGAT

AAGTCTCACCTAAATTCGCTGAGGTGGCACAAATTGGAATTCTTAATGTTCC**CAAGAGAAGAT**CAGGTGT
 CTGGACGAGCGGCTAGGAG**GAATGG**CTATTGGGGAAGCAGTTGAACAGGGGATTGTGAACAATGAGACACT
 TGGTTATTTTCATCGGGCGTGCGTATTTATTCTTGACCCAACTTGGGATTGACAAAGACCGATTAAAGGTTT
 CGCCAGCATCTGGCAAAT**AGATGG**CACACTATGCTGCGGATTGTTGGGATGCTGAAATC**GAATG**CTCTT
ATGGTTGGATCGAGTGTGT**TGGT**ATTGCTGATAGATCAGCATATGATCTGCGTGCCCACTGATAAAAG
 TGGAGTTCCCCTTGTGGCACATGAGAA**ATTCA**CAGAACCTAGAGAAGTCGAGAAGCTTGTATAGCTCCA
 GTTAAGAAAGAGCTGGGCCTTGCCTTTAAGGGTAGCCAA**AAGATG****T**GGTTGAAGCTTTGGAG**GCCATGG**
ATGAAAAGGA**GCCAT**GGAGTTGAAAGCAACTCTGGAAGCCAAAGGAGAGGCCTCGTTTCGTGTTTGCAC
 TCTTGAAGAACCGTGAACATAAAGAGTAAT**ATGGT**CTCCATCTC**CAAGA**AGAAGAAAAAGGAGCACCAG
 AGGGTTTTCTACTCCATCGGTTATCGA**GCCAT**CGTTCGGTATTGGGAGGATCATATATTGCCCTCATGAGC
 ACTCGTTTTACACGAGGCCTAGCAG**AGATGGAGATGAA**CAGCTCAATGTCTTTGGCTTCCCGCCTCTGT
 AGCCCCCATCAAATGCACTGTTTTTCTCTGGTTCA**GAATG**AGCAATTTGAGGAAGTTGCTAGAAGGAT**T**
GCCAGGTCATA**ACTGCC**GTTGGGATCACGTACAAAATCGACAT**TACTG**GAACCTCGATTGGGAAGAGAT
 ATG**CAAGA**ACG**GATGAA**CTGGGAGTCCCGTTAGCAATCACTGTTGACTCAGAAAGCTCGGTGACTGTGAG
 AGAGAGGGACAGCAAGCAGCAGATTTCGTGTTAGTGTT**GATGA**GGTGGCATCTGTGGTGAAGGAGTTT**ACA**
GATGGAGTGAGCACCTGGGGCGACGTGCTATGGAGATACCCCACT**CATTCAA**ATTGT**GATGA**CGTGTCGT
 GGATTGTCAAAAACCTTTGCT**ATGAAT**CAGATCGCAGATATGAGGCCCAATATCGTATTCC**TCAAG**GCCT
 TCTCTCGCT**TACTG**AAATTTCTTTCCGATAAGGTCATCCCGGACATCAACTCACCGGCCAGACAT**CTGCC**G
 CCGGCACTCGGGCCGCCAACGGTTGTGAAAAGC**ATGGTA**GTTTTGAAAAGAGAAGTATTTTATCTGAGAT
 CATCATTTTCGTGCGAGCCAA**ACTGC**AGTTTTCTGCACAAAACCTGAGCGCACTTTTGCTTCATCGCAAGC
 TCTTTCCGTAAC**AGATGAT**CAAACACTTTTT**CAGCATTCTTGAAT**CTTGCAAGCTGTCCCACAATTT**CAGA**
ACTGCAAT**TACTGT**GCACAGTAGAATCATTAACATGGGTATG**GAATG**TAC**CCATT**GCTATTATCCACGT
TACTGAAGGTGTACATGGCT**TGCC**AAAAAC**CCATTCT**TGCTCAACAACTACTTGCCGAGGTATGTTGTCC
 AAATTTTGGCATCTCTGCAAATTTGCTGATTGATAGTTGTTTGAAGACTGGCGAAATAGATATGGCAGAG
 AAAGTATTTTCTGCTCTACCTATCCG**AGATG**TCGTCACATGGAACCTGATCATTGGGGGTAATGTCAAAA
 GTGGATTCTTTGTGGAAGCAATTGAGTTGTTTAAACAAATGTTAGAAACAAATATTGAGCC**AGATG**GGTT
 TACATTTGCTTCAGTCATAAACGCGTGCT**TCAAG**GCCTTGAGCTCTTA**GCCAT**GGAAAGTGGGTGCATAGT
 TT**GATGAT**CGAGAAAAAATCGAGTTAAATTATATATACTTTCTTCTGCACTTATAGACAT**GTACTCTAGAT**
GTGGTAGAAATTGAAACAGCAAAATCCATCTTTAGCAATGCTCAGAAAACGGATGTATCCATATG**GAATG**C
 AATTATA**AATGG**ACTAGCA**ATTCA**TGGGCTTGCTGTGGATGCAATTTCTGTTTTCTCA**GCCAT**GGAAAAGG
 GAAA**AAGAT**ATTTACCTGATTCTGTAACCTTTATGGGAATCATAAAGCTTGATGCC**ACTGC**GGCTTAC
 TAGAACAGGGTCA**AATGGT**ATTTTAATCTAATGCAAAAAGTTACTTGATTGCGCCGAAATCGAGCACTA
 TGGGGC**AATGGT**CGATCTTTTTCGCACGAGCCGACTTCTGGAAGAAGCTTATGCAATGATTAAT**AGATG**
CCAATGGAGCCTGATGTTTTTGTATGGAGGACACTACTTAGTGATGTAGAACTCACAAAAAATCCAAGT
 TGGGTGAAGAAGTTGTGCGGAAAATAAAGCATTTAGGCAGTGGGGATTATGTTTTGATGTCAAATAT**GTA**
CTGCTCAACAAAGA**AATGG**GATTATGCAGAAAATATGAGGCCAACT**ATGAAT**AAGAAGGGAGTTCATAAG
 GCTAAGGGAAAGAGTTGGGTTGA**ACTGCATGGTATGATTCA****CCATT**TTAGGGCAGGTGATAGGTCCCAACC
 CAGAAACCGAATCAATTTATACACTATTGGAAGCTTTGACTCGTCGATCTA**GAATGG**AAGGTTATGTTTC
 TGTACAGGATTTGGTTTTGATGGATATATCAGAAGAAGAAAAGGAGGAGAACT**TGAAT**TATCACAGTGAA
 AAGTTGGCTGTGGCTTGCGGGATCCTGAAATCTAGTCCTGGAAGTGAATATTGATTTCCAAAAATTTGC
 GAACTTGTCTTGATTGTACAGTTGGAT**TCAAGAT**TGTGTGCAAGTTCTGCATAGGGTGATAACATGAG
 GGACAGGATTTCGCTTTATAGGTTTGAGAAGGGCTCTTGTTCTTGTGGTGATTAT**TGGTAA**

62 1.6% 3768 29 (5.3) 6.1 10⁻⁹

Sugiyamaella lignohabitans glycine--tRNA ligase (GRS1), partial mRNA NCBI Reference Sequence: XM_018879933.1

ATGTCCGAATCAACTATTACTTCCAGACCCCCAGTTCAGGATCGAGAAGTCCT**TGAAT**CTCTGCTGAAAC
 ACAGATTTTTCTACGGTCCTTCTTTTCGAAATATACGGCGGTGTTTCTGGATTGTACGATTACGGTCCAGC
 TGGTTGCGCCTTACAAGCCAACATTATCGACACTTGGCGTAAACACTTCGTGCTTGAAG**AAGAT**ATGCTT
 GAGGTCGATTGTCTCTGCTTTGACTCCCTACGATGTTCTTAAGACTTCTGGTCACGTTGATAAAATTTGCCG
 ATTGGGTTTCTCGTGATTTGGCT**TACTGG**TGAAATCTTCCGTGCGGATCACTTGGTCGAGGAAGTGCTTGA
 AAGCCGCC**TCAAG**GGTGATAAGGAAGCTCGTGGTGTCTGAGAATCAACCGACC**CTGCCA**AGGAAGAAAAA
 GACAAAAAGCG**CAAGA**AGAAGGTTAAGGAGACCAAGGCTATCAAACCT**GATGA****TGAA**ACTGTTGCTGAGT
 ATGAGTCAATCTTGGCCCAAATTGATGGCTACTCTGGCCCTCAATTGGGTGAGTTG**ATGGTA**AAGCA**CAA**
GATTGTTAATCCTACCACTGGTGGAGAATTGGAAGCTCCAAGGGA**ATTCAA**TCTTATGTTGAGAC**GCC**
ATTGGCCCCCTCTGGTCAATTGAAAGGTTATCTTAGGCCTGAA**ACTGCC**CAGGGTCAATTTTGAACCTCT
 CCAAACACTTGAAAGTAACAACCTC**CAAGATGCCATT**CGCATCAGCTTCTAT**TGGTA**AATCATTTAGAAA
 CGAGATTTCTCCTCGTGCTGGTTTTACTTCGTGTTTCGAGAATTTTTGATGGCCGAAGTCGAACATTTTCGTC
 GACCCTAACGATAAATCTCATGCTCGTTTT**GATGAA**ATTAAGGATATTCGATTGAGATTATGGCCCATGG
 ATG**TTCAAGCTCAAG**GTAAGTGAATATCAGCGAACTTTCTATCGGCGAAGCTGT**TTCAA****CTGGTA**TGAT

TGACAACCAAACCTCTCGGTTATTTCCCTTGCTCGAATTTACCTCTTCATGATT**AAGAT**CGGTATCGATCCC
AAGAGAATGAGATTTAGACAACACATGGCCAATG**AGATG**GCACACTATGCTGCTGATTGTTGGGATTGTG
AACTGCAATCTTCTT**ATGGT**TGGATTGAGTGTGTTGGATGTGCTGACAGATCTGCTTATGACTTGACTGT
 GCACT**CTGCCC****GTA**CTAAAGAAAAGCTCGTTGTCCGTCAAACCCCTCGACGAACCGGTTACCGT**TGAAT**CA
 TTTGAGGTGGAGTTGCAAAAGAAGAAGTTTGGTCCTAAGTTCCGTAAGGATGCTCCTAAGGTGGAAGAGT
 TCTTGCTTAGCCGTAC**GAATG**TGAACGTGAGAACCTCGCC**CAAGA**ACTCAAAGACACGGGTAAAATCGT
 TTTTGAGGTGGCCGAATTGATGG**CAAGAT**CGAAGTTGACTCATCATTTATCACTATTGAAAAAC**GTA**CT
 AGAGTTGAACATATTAGAGAGTACACTCCTAATGTAATTGAACCCCTCATT**TGGTA**TTGGACGTATTCTTT
 ACTCACTAATTGAACACTCTTTCTGGACTCGTCCCG**AAGATG**CTCAACGTGGTGTGTTATCCTTCCCACC
 AACTGTTTCTCCTACCAAGGTTCTTGTCGTCCCCCTTTCCAGCAATGCAG**ATTCA**CTCCCATCGCTCGT
 GAACTGGTTTCCAAGCTTAGACGTCTTG**GCA**TT**CAAT**TCCAAGGTCGACGACTCATCCGCTCTAT**TGGTA**
 AGCGTTACGCTCGTAATGACGAGCT**TGGTACT**CCTTTCCGTTATTACTATTGATTTCCAAACCG**TCAAGGA**
TGGTTCTTTAACTATTTCGTGAACGT**GATTCA**ACCAAGCAAGTCCGTGGCAG**TCAAGATGAA**GTGCTTGAG
 GCACTCCGCTCTTTAATTTTT**ATGGT**GTTTCCTGGGAGTCGGGAACATCACACCTTGCTACATTTGTCT
TCAAGATATTGAGTAG

38 1.9% 1973 15.4 5.8σ 3 10⁻⁹

Saccharomyces cerevisiae S288C glycine--tRNA ligase (GRS1), partial mRNA NCBI Reference Sequence: NM_001178469.2

ATGAGTGTAG**AAGATATCAAGA**AGGCTAGAGCCGCTGTT**CCATT**TAACAGAGAACAGCTAGAAAAGTGTTT
 TGAGAGG**AAGAT**TTTTTTTATGCCCCAGCTTTTGACTTGACGGTGGGGTTTCTGGTTTATATGACT**ATGG**
TCCACCAGGTTGTG**CATTCC**AAAACAACATCATTGACGCCTGGAGAAAGCATTTTATTTTGAAG**AAGAT**
 ATGTTGGAAGTTGATT**GTACT**ATGTTAACTCCATATGAGGTTTTGAAGACCTCAGGTCATGTTGACAAAT
 TTTCTGATTGGATGTGTAGAGATTTGAAGACTGGTGAGATTTTCAGAGCTGA**CCATT**TAGTCGAAGAAGT
 TTTGGAAGCTCGCTTAAAGGGTGAC**CAAGA**AGCCAGAGGTTTAGTCGAGGATGCTAATGCCGCTGCTAAG
 GACGATGCCGAAAAGAAGAAGAGAAAGAAAAAAGTTAAACAAATTAAAGCAGTCAAATTG**GATGA**CGACG
 TTGTCAAAGAAT**ATGA**AGAAATTTTGCTAAAAT**AGATG**GTATTCCGGCCCCAGAATTAGG**TGAAT**TGAT
 GGAAAAATACGATATCGGTAACCCAGTCACCGGAGAACTTTAGAATCTCCAAGGGCTTTTAACTGATG
 TTTGAAACGGCTATTGGCCCATCTGGTCAATTAAAGGGTACTTGAGGCCAGAA**ACTGCC**CAAGGTCAAT
 TCT**TGAAT**TTTAAACAAGTTGTTAGAATTTAAACAACAGTAAGACCCCTTTTGCCCTCAGCCTCTATCGGTAA
 ATCATTTAGAAATGAGATTTCCCCAAGGGCCGGTTTATTAAGAGTCCG**TGAAT**TTTGTATGGCAGAGATT
 GAACATTTGTTGAC**CCATT**GGACAAATCTCATCCAAATTTA**ATGAA**ATCAAAGACA**TCAAG**CTTTCCT
 TTTTGCCCTCGTGATG**TTCAAGA**AGCTGGTTC**TACTGA**ACCAATTG**TCAAGA**CAGTTGGTGAAGCAGTTGC
TTCAAGAATGGTCGATA**ATGAA**ACTTTGGGTACTTTAT**TGCCA**GAATTTACCAATTTT**GATGAA**AATT
 GGTGTT**GATGA**GTCCAAATTGAGATTCCGTCAACATATGGCTA**ATGAAATGG**CACATTATGCTGCTGACT
 GTTGGG**ATGGTGAAT**TGAAGACATCTTACGGGTGGAT**TGAATG**TGTCGGATGTGCTGATAGAT**CTGCC**TA
 TGATTTGACCGTT**CATTCC**AAAAAGACCAAAGAAAAATTGGTTGTTAGACAAAAATTAGATAATCCAATT
 GAAGTCACCAAGTGGGAAATTGATTTGACGAAAAAATTGTTCCGGTCCAAA**ATTCAGAAAGATG**CACCAA
 AAGT**TGAAT**CTCATCTATT**TGAAT**ATGTCT**TCAAGATGA**CTTGCGCTTCCAAGGCTGAGTTATTGAAGGCAA**AA**
TGGCAAGTTTACTATAAAGGTT**ATGGT**GTTGATGGCGAAGTCGAGTTGGACGATAAATTGGTTAAATC
 GAACAAAGAACAAGGTTGAACACGTTAGAGAATATGTTCCCTAGTGTCATTGAACCATCTTTCGGTATTG
 GCCGTATCATCTACTCCGTTTTTGAGCACTCTTTCTGGAACAGACCAGAAGACAA**TGCCA**GAAAGTGTTCT
 TTCCTTCCACCTTTGGTTGCTCCAACAAAGGTTCTTTTAGTTTCCTTTATCTAACCACA**AAGAT**TTAGTT
 CCTGTTAC**TACTGA**AGTCGCCAAAATCTTAAGAAAGTCTCAAATC**CCATTAAAGATGATGA**TTCTGGAG
 TTT**CCATTGGTAAAGAT**ATGCCCGTAATGACGAATTA**GGTACTCCATT**CGGTGTCA**CCATT**GACTT**TGA**
ATCTGCCAAGAC**CCATTCT**GTCACTTTGAGAGAAAGAGATTCCACCAAGCAAGTTAGAGGTTCCGTCGAA
 AATGTCA**TCAAG**GCTATCCGCGATATCACTTAC**AATGGT**GCTTCTTGGAAGAA**GGTAC**CAAGGATTTGA
 CTCCTTTTATTGCCCAAGCTGAAGCTGAAGCTGAAACTGACTAA

38 1.9% 2000 15.6 5.6σ 10⁻⁸

Archaeoglobus veneficus SNP6, complete genome GenBank: CP002588.1: c173853-172147 Glycyl-tRNA synthetase

ATGAGCGAAATCATGGAAATGCTCATCAGAAGGGGCTTTCTCTGGCAATCCTTCGAAATATACGGAG**GAA**
TGGCCGGATTTATTGATT**ATGGT**CCTCTCGGT**AATGG**ATTGCGTAGGAGGCTCGAGAACCTCTGGCGCGA
 GTTCTTCGTAGTTAACGAGAGAGCTGTGCGAGATGACACGCCAACCGTAGGAATAGAGGAAGTTTTCATA
 GCCTCCGGCCACGCCACATCCTTCACCGAGCTGGCTAT**TGAATGCAAGAA**GTGCGGTAGAGTTTTCAGGG
 CCGACCACTACATAAAGGAGAAGCTCGGAATCGAGGCT**GATGA**GACCGTTGAGGCGGTGAAGGAGATCCT
 CGAGAGCTTCGACTTAAAGTG**TGAATG**TGACGGTGAGTTTGGAGAACCAACGCCG**TTCAA**CCCTCATGTTC

GAAACGAAAATAGGGCCGGGGAAGGGGAAGAAGGGCTACCTGCGCCCTGAAACGGCCCCAGGGAAT**ATTCA**
 TTGCC**TTCAAG**AGGCTGGCGGACTATTTTAGAGATAAGCTTCCCTTTGGAGTTGCCCAGATTGGAAGGGC
 TTACAGAAACGAGATAAGCCCGAGACAGGGTGTGATCAGGCTCAGAGAG**TTCAA**TCAGGCCGAACTTGAG
 TTCTTCGTCCATCCGGGCGAGAAGAAGCATCCTGATTTTACATTTATGC**AGATGATGAAGTCAAG**CTTC
 TCGACAAGTTTGGGAACGAGCATGTCATAACTCTCAAAGAGGCTGTTGAGAAGGGGAATAATAGCCAATGA
 GGTATATAGC**GTA**CTTCATTGGAAAAACGAGGAGATTCTTCTTCAGGCAGGTATAGAC**GATGAA**AGGTTG
 AGGTTCCGGCAGCACCCGCGAC**GATGAA**AGGGCTCACTACGCCTCGG**ACTGCT**TGGGATGCAGAAGCCTTAA
 CGAGCTACGGCTGGATAGAGATCGTAGGAATTGCAGACAGAGGCGACTACGACCTCAGCAGACAC**ATGAA**
 GTACAGCGGAGAGGACTTGAGTGTCTTCGTACCCCTTTAAAGAGCCCGTAAAGGTGAAGAGGAGA**AAGAT**C
 GTTCCAAACATGGGCAAGCTCGGGCCGA**CATT**CAGGCAGAAGGCGAAG**AAGAT**TGCCGAAGTTCTTGAAA
 GCATGGAGATTGCAGACGATTCTATTGAG**GCCAT**CGAAGTAGAAGT**AGATG**GAGAGAAAAATATCAGTGAG
 CAGGGAGTTTTTCGAGGTCAAAGAGGTGGAGGAGCTCATCCACGGTGAGAAGGTTACTCCACACGTCATA
 GAGCCCTCTTTCGGTCTCGACAGGATTACCTTTGCTATCCTCGAGCATGCGTTGCACAAAAGACATCGTTG
 ACGGGGAGGAAAGAAGAGTTCTGAGGCTCAAAGGTGGATGGCTCCCGTTCAGGTTGCTGTG**CTGCC**GCT
 TCTCTCAAAGGTGCTTTTATCGAGAAGTCCATCGAGATAA**CATT**CATGCTTAAGAAAGAGGGTATTTTC
 ACAGAATACGACGACTCCGGAAGCATAGGAAGGA**GGTAC**AG**AAGAT**TCGACGAAATC**GGTAC**GCCTTTCT
 GCGTTACGGTTGACCACGACTTTTGAGGACGAGACAGTAACGATAAGGGACAGAGATACAACGAGACA
 GATAAGAGTCCATATAGGGGGTCTCGCTCCAGCACTTAGGGAGCTGCTCAGTACAGACA**AAGAT**ATATCC
 GAGTTTGAGAGGTCTTTAAAGAATAG

23 1.35% 1703 13.3 (3.6) 2.7 10⁻⁵

Table S6. Nucleolin sequences from different species with after the sequence: Nb of observed pentamers among the set {ATTCA TTCAA TCAAG CAAGA AGATG GATGA ATGAA TGAAT} with percentage among the set of pentamers of the sequence (in red), Nb of nucleotides in the sequence (in black), Nb of expected pentamers with its standard deviation (in red), Nb of standard deviations between observed (in blue) and probability of this event (in green).

Bauhinia variegata isolate BV-YZ2020 chromosome 1, whole genome shotgun sequence GenBank: JAKRYI020000001.1: 6791445-6791571,6791692-6791734,6791876-6792039,6792161-6792274,6792411-6792530,6792613-6792729,6792815-6792922,6793010-6793024,6793315-6793389,6793503-6793604,6793720-6793824,6793948-6794052,6794161-6794277,6794388-6794507,6794592-6794672,6794760-6794851,6794924-6795041,6795250-6795338,6795438-6795528,6795666-6795757,6796366-6796655,6796872-6797144 nucleolin

TCCCCAAACCCCTAAACTCTGTGACGCGTAGGGTTTTGCAGAGCTTCTCGTTATCTAT**CCATT**TCCTCGAC
 CTTTTTCTATTCTTTGTGTCTAATATGGCAAAGTCTAG**CAAGA**AATCAGCTTCCAAAGTTGAGTCCGCTC
 CAGCCGTCGTTCCAAAATCTGCGAAGAAAGGTGCTAAGCGAGAAGCTGAG**GATGA**GCTTGAGAAGCTGGC
 GAGTGCAAAGAAGCAGAAAAG**AGATGA**GCTTGCTGACAAGCAAT**TGAATG**CGAAGGTGCAAAAAGAAAAAG
 AAACCGCAAAGTAGTTTCATCTGATTCTCTTCATCCAATTCTGAGGAGGATAAGAAACCTGCTGTTAAAG
 TTCCTATTCC**TTCAA**AGAAGCAACCTGGCCCTGCAAAGAAAAGCAAGCCAGCAAGCAGCTCCAGTTCTGA
 CTCTTCT**GATGATCAAGT**GAG**GATGAAG**CTCCAAAATCTAAGG**TGAATGGAC****CTGCC**GCGGGTAA**GAAT**
GGACCTGTCTGCTACACGAAGAAGTTGATCTAGTGAAGCTCTGACTCT**GATGAC**CAGCAGCTCG**GATG**
AAGATAATGCTCCAAAATCCAGGGTGAACGGAGCGGCTGCT**TGGTAAGAATGG**ACCTGCTACTACAAAGAA
 AAAAGCTGATTCTAGTGAAAGCTCTG**ATT**CAGATGACAGCAGTT**AGATGAA****GAT**TTGGCAGCTGCAGTT
 GCTCCTAAGAAGAAAGTGGAAGCAGTTGCTCCAAAGAAAAAGTTGAGAGCAGTGACAGCTCAG**AAGATG**
AATCAAGCGCTGAAG**AAGATAATGATTCAA**AACTAAGGAA**AAGATG**GATGTT**GATGATGA**TAGTTCTGA
 AGAAAGTGAAGAAGAGCCACAAAGAAAAATATCAAAGGCT**TGCCA**AATGCTGGAAAACCTAATGTAAAA
 GCTGTCAAAAAGGATAGCAGTAGTG**ATGGT**TCAGAG**GATGATGA**TAGCTCTGAAGAG**GATGA**CAAAGTAA
 AAAAGCCAAATGCTGGAAAACCTAATGTAAAGCTGTTGAAAAGGGTAATAGTAGTGATAGCTC**AGATGA**
TGAAGATGATAGCTCTGAAGAG**GATGA**GAAAGTGAAAAAGCCAAATGCTGCAAAACCAAGTGTAATCT
 GTCAAAAGGATAGCAGCAGTGATAGCTC**AGATGATGAAGATGA**TAACCTCTGAAGAG**GATGA**GAGAGTGA
 AAAAGCCGAT**TGCCACATCAAG**TGCTGGAAAACCTGGTGTAAGTTGTCAAAAAGGATAGCAGTAGTGA
 TAGCTC**AGATGAAGATGA**TAGCTCTGAAGAGGAG**GATGA**GATAGTGAAAAAGCCAAAGACTACACCAAAT
 GCTGGAAAACCTAGTGCAAAATCTGTCAAAAAGGATAGCAGTAGTGATAGCTC**AGATGATGAAGATGA**TA
 GCTCTGAGGAGGATGTAAATCTGTCAAAAAGGCTA**AAGATGTGAAGATG****GT**TGATGCAGCCTCACCAGA
 GCCAACTT**CAAGA**ATTTTA**ATT**CAGCATCTGCTAAGAAAACCTCCTAAGACTCCAGCTACCCCTGTAGAA
 CAAAATTCACAGGAAGGACACTATTTGTTGGAACTTGTCTATATAGTGTGGAAAGAGCTGATGTGGAAA
 ATTTCTTCATAGATTGTGGAGAAGTTGTTGATGTCCGCCTTGCTGT**AGATGATGAA**GGAAGGTTTAAAGG
 TTTTGACAGATGTTGAGTTTGCT**TACTGC**AGAAGCAGCCCAAAATGCCCTGAAGT**TGAAT**CAAA**ATGAAT**TG
 TTGAGACGACCTGCAAACTTGATCTGGCAGCAGAAAGGCCCTCATATACCCCTAACAGCAGCAACTGGA
 GCAACTCCTTCCAGAAAGGTGGAAGAGGTGAGTCTCAAACAGTATTTGTTAGGGGTTTTGACTCATCCCT
AGATGAACAGCAGATTAGGTCTAATCTGGAGGAGCATTTTAGTTCTGTGGAGAAATTACTAGGATATCA
 ATCCCTA**AAGAT**TATGACACTGGTGCTTTTAAAGGGGTATGCATACATGGACTTGAGTGATGCT**AATGG**CA
 TAAACAAAGCATTAGAATCAATCAATCAGAGTTAGGCGGTTAT**CCATTGATGGT**TGAAGAAGCAAGGCC

78 3% 2554 20 (4.47) 13 σ 10⁻³⁰

TTGATACTTCTTTATCAGTCTTTAGCATTTTCAGTAAGTGCTTCTGAAGGCAGTGT**TTCAAA**CTAGTAT**TTCAA**ATTCTGACGAATCAGATAGTGCTAGTGAAA**GTA**CT**TC**AAATAGTACC**TC**AAGTAGTATATCTAATTC
TGACGCTTCTACTAGTTCCTAGTGAGAGTACCTCAACTAGTTCATCTATAAGTGC**TTCAAA**CGAATAAAAGT
GAAAGTGACTCAATCTCTGAAAGTGGATCTGGAACAAACTCACAAA**GTA**CTTCTCAATCAGC**TTCAAA**CAA
GTACAAGTGAAAGTG**ATTCAAG**CAATAGCGAAT**TC**AAGTTCTGTGACGCATAGTGACTCAGCCTCCACAA
TGCATCTGAGAGTGCATCAGCATCAGATAG**GTA**CTAGCA**ATTCA**GCAGCGAAATCGGAAAGTAACTCGGCA
GCTAGCGATAG**TGAAT**CAACGAGTGCT**TTCAAG**TTCTAATGCTTCTACAAGTGCTAGTGATAGTGCT**TTCAA**
CGAGTG**GATTCA**ACTAGCACCTCAACCA**GTACTTTCAAG**CTCCAATGCCTCGAAGAGTGCGAGCGACAGTGCT
TTCCATAAGTACCTCTATCAGTGATTCTAC**CAAGA**AGAGTGATAGTGACTCAGCTTCTATGAGTGGCTCA
GGTACAG**ATTCA**GAAAG**TGCCA**GCAAGTCAGCTTCCACGAGTGCGAGTGCTAGTGATTCTGCAGC**CAAGA**
GTACAAGTG**ATTCA**GTAAG**GTA**CTTCAGCATCTGGAAAAGACAGT**AATGG**CTCAACTAGTTTCATCAGCTTC
AGCCTCAACGAGTTTCATCTGAAGC**CAAGA**GTACAAGTGATAG**TGAAT**CCACTGTAAATAGTGGATCAGAA
TCAGCAA**GTA**CTTCAGCT**TTCAAG**CTCTGACAGTGCT**TTCAAG**CAGTAAGTCAGCATCAGACCATGGGTCT**TG**
AATCAACGAC**TGAAT**CGGCTAGTGCT**TTCA**ACCAGTATTTTCAGATAGCGAAA**GTA**CTTCGAAGAG**TGAAT**C
AGCAAGTGCT**TTCA**ACTAGTAGAGCATCACAGTCAGATAG**TGAAT**CACAAGCTAGTGAAAGTGATACAAGC
AGCAG**TGAATCAAG**TTCTGTGACGCATAGTGCGTCAGCTTCCACAAGTGATCTGATAGTGCATCAGCAT
CAGATAG**GTA**CTAGCAATTCCGCAGCAAAAATCGGAAAGTAACTCCGAAGCTAGTGATAGTGCT**TTCAA**CGAG
TGCTTCTAGTTTCCAATGCATCGAAGAG**TGCCA**GCCT**TTCAA**CTAGTTTCTCAGCT**TTCAA**CTTCCACGAGT
GCTAGTGCCCTCAGTTTCTGATAGTGTTCCTAATAG**TGAATTAATGAGTACTTCAAG**TTCTAATG**ATTCAA**
TTCAAATTATTGTTTCTGATAGTGCTTCCACTAGTAATTTAAG**TTCAA**ATACTACTAGTACAAACACTAA
CGAATCAGAATCAAATCTTCTGCAACTATTGCTAGTAGTGAGGAGTAGTAATAACTCTTGCAACTCAATCA
GCTCAAA**ATTCA**TCAGCA**AATGGTAATGAAT**TGAACTCAACAAGTTCTAACAAGTCTGCTAAGAATCAAA
AGACTAATTTACCACAAACTGGA**CAAGATTCA**TCAACTGGATTTATGAGTTCTATTGGTTCTATGATTAT
TTCAATCATGGCTCTCTTAGGCCTAGGAAGACTAAGAAAAGAAAAAGG**AAAGAT**AAATAA

55 3.3% 1664 13 (3.6) 11.7 σ 3 10⁻²⁸

AGTCTCGAGCTCTCGCTGGCCTTCGGGTGTACGTGCTCCGGGATCTTCAGCACCCGCGGCC**GCCAT**CGCC
GTCGCTTGGCTTCTTCTGGACTCATCTGCGCCACTTGTCCGCTTCACACTCCGCC**GCCAT**ATGGTGAAG
CTCGCGAAGGCAGGTAAGAA**TCAAG**GTGACCC**CAAGA**AATGGCTCCTCCTCCAAAGGAGGTAGAAG**AAG**
ATAGTG**AAGATGA**GGAAATGTCAG**AAGATGA****AAGATGA**TAGCAGTGGAGAAGAGGTCGTCATACCTCA
GAAGAAAGG**CAAGA**AGGCTGCTGCAACCTCAGCAAAGAAGGTGGTCGTTTCCCCAACAAAAAGGTTGCA
GT**TGCCA**CACCAGC**CAAGA**AAGCAGCTGTCACTCCAGGCAAAAAGGCAGCAGCAACAC**CTGCCA****AGA**AGA 23
CAGTTACACCAGCCAAAGCAGTTACCACACCTGG**CAAGA**AGGGAGCCACACCAGGCAAAGCAT**TGGTA**GC
AACTCC**TGGTA**AGAAGGGTG**CTGCCAT**CCCAGCCAAGGGGGCAA**GAATGG****CAAGA****TGCCA****AGA**AGGAA
GACAGT**GATGA**AGGAG**GATGATGA**CAGTGAGGAG**GATGA**GGAG**GATGA**CGAGGACGAG**GATGA**GGATG
AAGATGAATTGAACCAGCAGC**GATGA**AGCAGCAGCTG**CTGCC****CTGCC**TCAGAG**GATGA**GGAC**GATGA**
G**GATGA**CG**AAGATGATGA****GATGA****GATGA****GATGA**GG**AAGATGA**CTCTGAAGAAGAAGCTATGGAGACT
ACACCAGCCAAAGGAAAGAAAGCTGCAAAAGTTGTTCCGTGTAAAGC**CAAGA**ACGTGGCTGAG**GATGA**AG
ATGAGAAGAG**GATGATGA**GGACGAG**GATGA**CGACGACGACC**AAGATGATGA****AAGATGATGATGATGA****GA**
TGATGAGGAGGAGGAAGAAGAGGAGGAGGAAGAGCCTGTCAAAGAAGCACCCTGGAAAACGAAAGAAGGA**A**
ATGGCCAAACAGAAAGCAGCTCCTGAAGC**CAAGA**AACAGAAAGTGGAAGGCACAGAACCAGCTACGGCT**T**
TCAATCTCTTTGTTGGAAACCTAAACCTTAAACAACTCTGCTCC**TGAAT**TAAAAAC**TGGTA**TCAGCGATGT
TTTTGCTAAAAATGATCTTGCTGTTGTGGATGTAGCAAT**TGGTA**TGACTAGGAATTTGTTTATGTGGAT
TT**TGAAT**CTGCTGAAGACCTGGAGAAAGCGTTGGAACTCACTGGTTTGAAGTCTTTGGCA**ATGA**ATTAA
AACTAGAGAAACCAAAAGGAAAAAGACAGTAAGAAAAGAGCG**AGATG**CGAGAACACTTTTGGCTAAAAATCT

CCCTTACAAAGTCACTCAG**GATGAAT**TGAAAGAAGTGTTT**AAGATG**CTGCGGAGATCAGATTAGTCAGC
AAGGATGGGAAAAGTAAAGGGATTGCTTATAT**TGAAT**TTAAGACAGAAGCTGATGCAGAGAAAACCTTTG 80
AAGAAAAGCAGGGAACAGAGATCGATGGGCGATCTATTTCCCT**GTACTA****TACTG**GAGAGAAAAGGTCAAAA
TCAAGACTATAGAGGTGGAAAGAATAGCACTTGGAGTGG**TGAAT**CAAAAACTCTGGTTTTAAGCAACCTC
TCCTACAGTGCAACAGAAGAACTCTTCAGGAAGTATTTGAGAAAGCAACTTTTATCAAAGTACCCCA
ACCAA**AATGG**CAAATCTAAAGGGTATGCATTTATAGAGTTTGCTT**CATT**C**GAAG**ACGCTAAAGAAGCTTT
AAATTCCTGTAATAAAAGGGAAATTGAGGGCAGAGCAATCAGGCTGGAGTTGCAAGGACCCAGGGGATCA
CCTAA**TGCCA**GAAGCC**AGCCAT**CCAAACTCTGTTTGTCAAAGGCCTGTCTGAGGATACCACTGAAGAGA
CATTAAAGGAGTCATTTGACGGCTCCGTTCCGGCAAGGATAGT**TACTG**ACCGGAAACTGGGTCTCCAA
AGGGTTTGGTTTTGTAGAC**TTCAAC**AGTGAGGAGGA**TGCCA**AG**CTGCCA**AGGAG**GCCAT**GGAAGACGGT
GAAATTGATGGAAATAAAGTTACCTTGAGCTGGGCCAAACCTAAGGGTGAAGGTGGCTTCGGGGGTCTG
GTGGAGGCAGAGGCGGCTTTGGAGGACGAGGTGG**TGGTA**GAGGAGGCCGAGGAGGATTTGGTGGCAGAGG
CCGGGGAGGCTTTGGAGGGCGAGGAGGCTTCCGAGGAGGCAGAGGAGGAGGAGGTGACCACAAGCCACAA
GGAAAGAAGACGAAGTT**TGAAT**AGCTTCTGTCCCTCTGCTTTCCCTTTT**CCATT**TGAAAGAAAGGACTCT
GGGGTTTTT**TACTG**TTACCTGATCAATGACAGAGCCTTCTGAGGA**CATT**C**CAAGA**CAGTATACAGTCCTGT
GGTCTCCTTGGAATCCGTCTAGTTAACAT**TTCAAG**GGCAATACCGTGTGGTTTTGACTGGAT**ATTCA**T
ATAAACTTTTTAAAGAGTTGAGTGATAGAGCTAACCTTATCTGTAAGTTT**TGAAT**TTATATTGTTTCAT
CCCATGTACAAA**CCATT**TTTTCTACAAATAGTTTGGGTTTTGTTGTTGTTTCTTTTTTTGTTTTGTT
TTTGTTTTTTTTTTTTTTCGTTTCGTGGGGTGTAAAGAAAAGAAAGCA**GAATG**TTTTATC**ATGGT**TTT
TGCTTCAGCGGCTTTAGGACAAATTAAGTCAACTCTGG**TGCCA**GACGTGTTACTTCCTAAAGAGTGTT
TCCCCT**GGAATG**TCACTGGAGAGCATGGCAAAGCCAGCT**CTGCCA**CTTGCTTCACCCATCCC**AATGGAAA**
TGGCTTAGTGCGTGTTCAGTATCCCAGCCCTAACTAACTTGGTTGAAATGCTGGTGAGGGGACCTGCT
CCTGCAGCCCTGGTGCTGACTTGAAGGCTGCTGCAGCTTCTCCTACTTTTAGCAGGTCTGAGGATTATGT
CCTGAAGACCACTCTGGAAAGAGGTGCAGGAACAGATTAGTCAGGTTTCTTAGGACAAGGAAGAGCTTCA
GGGAAGAGCAGTGGCTAACTCCTGTAATCCCAACACTTGGGGAGGCCGAGGCAGGCAGATCAACTGAGGT
CAGGAGTTGAAGCAGCCTGGCCAAC**ATGGT**GAAAGCCCATCTCTACTAAAAATACAAAAATAGCTGG
GC**ATGGT**GGT**GTACT**CCTGTAGTCCCAGCTACTCAGGAGCTGAAGCGGGAGAGTCACGTGAACCCGGGA
AGCAGAGTGAGCTGAGCACACACTACTATACTCCAGGCTGGGTAAACAAAGCGAGACTCCCATCTCCCAA
AAGCAGTTCTGGAATAGAATCACGCT**AGATG**GATAGACCAGTGACACTTTGGAACCTTGGGGCTGGGG
AGGAA**ACTGCC**CATCCAGTAAACCCCCAAAA**GCCATT**TGTTCTGCACTACGTATATTGCTTATTCTTTC
TGGTCTTAA**GTACT**TGCCTCTCAACCTCCCTTTTTAGTAAAGACAAGGCCACGTGAGAGGCGGGACTAT
CAACATTGT**GATGAAT**TTACTTGAAACCCAGTGCCCAAAATCAATGTAGGTAGCCAAGTCCAAAAACCTG
TTCTAGTCCAACCTAGTGAAATCAAACCTGTGATACTTGGATAAGCTTAGAAGGAAACGTGAAGAATACGTA
GCTGCTTTGGGTTTACTCTGGTTTCAGTTGGGCTGTTGAAATCTTAACATCCTTGGGCTTATCACC**TACTG**
CTTGTGAGCCCTGTTCCATGTCCAGGGGATGGGGGTGGTGACAATCCAGTT**CAAGA**CCCTCATGCTCTA
GAGAGGAAGGTGGCCAGCCAGGGTTGTAAC**TACGATGAAA**AGCAGTGGGAGGGTCTCCTATGAGGCAAG
CCTAAGGACAAAAAGGAAGGCCTTGCAGCCTGTATTCTGGATAAGGAATTAAGGCTCAGTTAATTGAAGCCCA

92 2.35% 3920 30.6 (5.5) 11.2σ 8 10⁻²⁶

PREDICTED: *Ornithorhynchus anatinus* nucleolin (NCL), mRNA NCBI Reference Sequence: XM_029064618.2

TCAGGCTCGAG**CTGCCA**GCAGGGAGCAGGAGCCCCGACCCGCGTCAAGCGCCCCGCCCTCCGCCTTTGCCCCGAG
CGATCTCT**CTGCC**CTTCCCGCTCGCCCTCTCTCCCTAGGCC**GCCAT****CATGGT**GAAGCTCGCGAAGGCAGCTAAAC
AAGGAGGACCGAAGAA**AATGG**CTCCACCTCCAAAGGAGAT**AGATGA**TAGTGAG**GATGAAGAT**ATGTTGGAG**GATG**
let7 3'-CTGTATGGTCT-5'
AAGAGGACAGCAGTGGAGAGGAGGTTGTTATCCCTCAGAAGAAGGCAGTGGCTACACCAGGGAAGAAGGGAGCTG
CCGCTGCCACCCCAGGGAAGAA**GCCAT**GGCCACCCAGGAAAGAAGGCTGTGGTGTCAACAGC**CAAGA**AGGCCG
TGTTGGCCACCC**CTGCCAAGA**AGGCAGCTGCAACACCGGG**CAAGA**AAGCAGTCGCGGCTGCAGCCACCCAGCC**CA**
AGAAAGCAGTACCTCTGCCAAAGCAGCACCCACCCAGGGAAAAAGGGGGCCACCCCGCCAAAGCAGCACCCA
CGCCAGGGAAGAAGGGAGCCACCCGGCCAAGGGGGC**CAAGA**ACGGAAAGAACGCC**CAAGA**AGGCTGAAAGTGAGG
AAG**AAGATGA**GGAAGAGAGTGAAGACAATGACGACTCTGAGACTGAAGCTTTGAAAATAGTTGCACCAACAAAC
ACGAC**GATGAATACGATGAGGATGATGATGATGACGATGACGATGATGAAGATGA**CTCTGAAGAAGAACCCATGG
AGATCACACCAGCGAAGGGAAAGAAGGCCCCACCTTTGAAGGCTGCTTTGAAAGCCAAGGTGGAGTCTGAGGAAG
AAGAGGAGGAG**BATGAAGATGATGGGGATGA**AGGAGGAGGAGGAGGAG**GATGA**AGGAGGAACCTGTGAAAGAAACAC
TTGGAAGAACGAAAGAAAGAA**AATGG**CCAAGCAGAAAGCTGCACCAGAGGCCAAAGAAACAGAAAACCTGAACCGATTG
AACCTGCATCAGCT**TTCAAT**CTCTTTATTGGAAACCT**TGAAT**TTCTCCAAA**ACTGCT**CCAGAGCTGAAAACAGGCC
TCAGTGATTTCTTTAC**CAAGAAT**ACCTCACCGTTCTGGATGTGAGAGTTGGTGTTCAGGAAATTTGGCTATG
TGGACTTT**TGAAT**CTGCT**GATGAT**CTGGAAAAGGCCCTGGAAT**TGAATGG**CTCTAAGGTCTT**TGTAATGAATGA**
AACTGGAGAAAGCAAAGGCCAAAGAGACAAAGAAAGAACG**AGATGCCA**AGGACCTCTTTGTGAAGAATCTTCCCT
ACAAAGTTAC**TCAAGAAGAGATGAA**AGAGGTGTTTGAGGATGCTGTTGATGTTAGATTAGTGACTGGCA**AAGATG**

GAATGAGTAAAGGGATTGCCTACATAGAATTTAAACAGAGGCTGATGCGGATAAGGCCTTGGAAGAGAAGCAGG
GAACTGAAATTGATGGACGTTCAATTATCCTGGACTACACAGGGGAGAAAAGCCAGGGCCTAGAAAACCTCCAGAG
GTGGAAAGAACAATGCCTGGAGCGGAGAGTCAAAGA**CCATT**GTCGTGAACAACCTTGCTTACAGTGCAACAGAAG
AAGGTCTCCAGGAAGTTTTTGAAGAGGCCAC**TTCAAT**CAAAAG**TGCCA**CAGAACACCAGGGCAGGCCTAAAGGAT
ATGCATTTGTAGAGTTCTCTACTCCTGAAGAAGCTAAAGAGGCTCTAAACTCCTTGAACAATGTAGAAATTGAAG
GTAGAACAATCAGACTGG**AGATG**CAAGGATCAAAAAATACTAACG**CAAGA**GGAGGGTTTCAGCCAC**CCATTCAAGA**
CTTTGTTTGTCAAAGGCCTGTCTGAAGAAACCACAGAAGAGACGTTGAAAGAGTCATTTGATGGCTCTGTTCTGTG
CTAGGATAGTCACTGACCGTGAAACTGGGTCTTCCAAAGGATTTGGCTTTGTGGAC**TTCAA**CACCGCAGAAGAAG
CCAAAGCAGCCAAAGAG**GCCATGGAAGATG**GAGAGATTGATGGAAATAAAGTTACGCTGGACTGGGCCAAGCCCA
AGTCCGAAGGTGGTTTTTGGAGGCCGCGTGGTGGCAGAGGAGGATTTGGTGGCAGAGGCGGTGGCCGGGCGGTG
GCCGAGGTGGGTTTCGGTGGCCGCGGCCGTGGAGGCTTTGGAGGACGGGGAGGCTTCCGAGGAGGCAGAGGCGGCA
GTGGTGGAGACAAGCAACCAGGAAAGAAGACAATTTGAATAGCTTCCCACTCCCCCTTTTCCCCCTCCCTTTCCC
TCTGCTATTTGAAGGACTCTAGGGTTTTTATTCTGTCTATCTTGTCACTGACGGAGCCTTCTGAGGA**CATTCCAA**
GACGACATGCAGTCCTGTTATCCACTTGAAACATAGATATCAT**TTCAAG**GGAAAGAGAGCGCGCTCTATTTTG
ACTGCTTGAGTATTCTGTATGGACTTTTTTAAAG**AGATGA**GAGGTAGATTGTGGGAGTTAACTTTGTCTGTGAGTT
GTCT**TGAAT**TTATAATGTTTTACCCCCCATGTACAAAAATTTTTTTCCT**ATGAAT**CTCTGTAGTTTTTTTTTGCT
GTAAAAACGC**GAATG**TTTTATC**ATGGT**TTTTGCTTCGGCAGCCTGTCTTGGGACAGATTAAAGGGCCTAAACTG
GTGG

70 2625 20.5 (4.5) 11σ 5 10⁻²⁵

PREDICTED: Acanthochromis polyacanthus nucleolin (LOC110950732), mRNA NCBI Reference Sequence: XM_022193488.2

GAGTAAGGAGCGCTCTTCGTCTAATCTGTCTTTCTGTCCAATGTCCCACCGAGTAGGCACAGCCACGTGCATGC
AGTAGCTGTTTTGTTACTTGTCTATAAACGTCCAGGTCCCCGTGTTTCACTGAAACACACCACAGCAGCACC**ATG**
GTGAAAGTTAGCAAAGGCAGCAAATAA**TCAAG**CAAGTCAAAAGAAAAGGCACCTCCACCCCTAAAGAAGTGGAG
GAGGAATCTAGTGAAGAGGACAGCAGC**GATGAG**GAAGAAGAGGGCCCCCTCCCCCAAAGTGGTGAAGAAAGCCACA
CCAGCTAAAGCAACCCCGTTGTTAA**AATGGTACTGCTGCCA**AAAAAGCAGAAAAGCG**AAGATGATGA**C**GATGA**C
GACGACGAGGA**AAGATGAT**TCTGAGGAGTCTTCTGAAGAGGAGGCTCCCCACCTAAGAAGACAC**CTGCCA**AGGCA
ACACCAGCTAAACCTGCAGCAGCTAAGGCAGCCCCCGCAAAGGCAGAGGAATCT**GATGATGA**CGAC**GATGATGA**
GATGATGATGAGTCTGAAGAAGAGGGCCCCACCACC**CAAGA**AAGCAGCCAAGCCAGCAGCTAAAC**CTGCC**GCCAAAG
GCCAAACCTGCAAAGGAGTCTTCT**GATGAAGATGATGATGA**CGAC**GATGA****GATGATGATTCA**GAAGAGGAGGCC
CCACCCT**CAAGA**AGGCTACTCCAGCTAAACAGCTGCAAAGCCTGCAGCAAAAGCCAAGCCTGCTAAAGAAGAG
TCGTCT**GATGAGGATGATGATGAAGAAGATGA**CTCTGAAGAGG**AGATG**GACACCCTC**CTGCC**CCAGCAGTC
GCCAAAGCAAAGAAAGCAGGC**ATGGT**GAAGGCCAAGGAAGAGTCAGAGGAGGAG**GATGATGATGATGA**CGACGAC
GATGATGATGAGGATGACGAAGAAG**AAGATGA**TGCCCCAGTGACACCTGGAAAGCGGAAGGCAGAGAC**CAAGA**AA
GACACACCTCCAGC**CAAGA**AAGCAAAGTCAGAAGGAGAAGGATTCTGTCTCTTTGTTGGAAACTTGAACCTCCAAC
AAAGACTTT**GATGAGAT**CAAAGCATCACTGAGAAAGTTCTTCAC**CAAGAATG**ACCTTGAGGTCTGTCGATGTCCGG
TTAGGAGGATCAAAAAAGTTTGTTATGTGGACTTTGGATCTGAAGAGGACCTGCAGAAGGCACTGGAGCTCAAC
GG**CAAGAAGAT**CATGGGTGAGGAGGTCAAACCTGGACAGAGCCCGAAGCAAAGAAGGCTCACAGGATGGCAGGAAA
GAGAGGGATGCACGGACACTGTTTGTAAGAACCTTCCCTTCTCTGCAACGGCT**GATGA**CCTGAAAGAAGTCTTT
GAGGATGCTGTTGACATCAGGCTACCTCAGGGCCAGAACGG**TTCAA**ACAGAGGCATTGCCTACATCGA**ATTCAA**
ACTGAGGCTGAAGCAGAGA**AAGATG**TTGGAGGAGGCCAGGGAGCTGATGTACAGGGAAGGTCTATAATGATTGAC
TTTGTTGGAGAAAAGAGCCAGAAAGGAGCCAAAGTGGCAGCATCAGGAGCAGTCAATAAAACCG**TGGTAGTGAAT**
AATCTTG**CATTCA**GTGCT**ACTG**AGGAAGTGTTCAGTCTACATTTGAGAAGGCAGTCTCTATCAGGATTCCACAG
AAAGATGGCAGGCCCAAAGGTTTTGCTTTTATAGAGTTTGAGAGCAAT**GATGA**TGCTAAGGAAGCTCTGGAAAAC
CTCAACAACGCAGACATTGAAGGCCGGTCAATCCGACTGGAGTACAGCCAGAACAGCGGT**CAAGA****GATG**GGGGC
AGGGGAAACTCGGGTCCAACAAAACCTCTTTG**TCAAG**GGTCTGTCTGAGGACACTACAGATCAGACCTTAAG
GATTCCTTTGAGGGTGTGTAGCAGCCAGGATAGTTACAGACAGAGACACAGGGTCCTCAAAAGGCTTTGGCTTT
GTGGACTTTGACAACGAG**GATGACTGCA**AGGCCGCGAAGGAGGC**AATGG**AGGATGGCGAGATTGATGGCAGCAAA
GTGACCTTGACTACGCCAAGCCCAAGGGTGAAGGCGCTTCCGTGGAGGCCGGGTGGCGGCTTTGGAGGCGGC
TTCGGTGGGCTGGTGGTGGCAGAGGAGGGCGGGCGGTTTTGGTGGTTCGCGGTGGCTTCGGCGGAAGAGGTGGA
AGAGGTGGTCTCGTGGAGGTGGACGCGGAGGCGGACGAGGTGGTTTTTGGAGGTGGAAGAGGTGGAGGATTTGGA
GCCAAACCCAGGGGAAGAAA**TCAAG**TTTT**GATGA**CTAAAGCCTTAGTCTGTTTTTTTTTTTTTTTT**TGAATGG**AC
TCTGGTTTTCATCAATTTTTCAGAGCTTTTGG**CATTCC**CAGAAAGCGTCAT**TGGTA**TATATTTAACTGTT**AATGG**GC
ATTTTAGCCCTTATTTAGACTCCTCAAACCTTCCACATCCTTCTG**TGCCA****GAATGGTACT**TTGACCTTTTCAGC
TTAAGTGCTATGTGTGAGTA**ATGAATGT**CCCC**GAATG**CAAACCTGCACCTCTTGCCCTTGACAGCTTACTTGCTTT
TCTGAAGGAGAAA**AATGGAATGGG**TCTGGCAAACAAGTTTGTGTG**AAGAT**ATGTTCTGTAGACATTGTGTAAA
TTTTAAACAGTTTGTATG**ATTCA**TTTTTACCTTTTGTAAAAAAGAAATACACCTGTATCCATGTTTTTTTGT
TTATTTTGTGTGTCATTTAATTTGCTTT**AGATG****CCA**AGTTT**TACTG**TAATGTGCAAAGCATATTTTGATAGCTTT
GAGCCCATGCTCAACAGAGTTGAGTGTTTAGGATTATGGATGTTTTTGTCTATTCTGTT**TGGTA**GGGCACAGTCAG

TGTTTAAGTTCAGGATGGAAGACCTAATCTATGTAACATAAAATCTTTGTAGTGCATTGACTAATTTTTCTCCCA
TCTCT**CAAGAT**TTTCTCGTAAATAAATAAAACCTTGTAGGTGGTGATTTTTTAAA

75 2.5% 2975 23.2 (4.8) 10.8σ 5 10⁻²⁴

Table S7. 5S of Archaea with their AL-pentamer observed (in red), Nb of nucleotides in the sequence (in black), Nb of expected pentamers with its standard deviation (in red) and Nb of standard deviations between observed (in blue).

Methanococcus voltae A00001 5srRNAdb

TGATACGGCGGTCATAGCGGAGGTGTCCCATCCGATC**CCATTC**CGATCTCGGAAATTAAGCCCTCCAGCGA
TTTCTTAA**GTA****CTG****CCATAT****G**GGGAA**CAAGATGA**CG**CTGCC**GATCAC

14 116 2.5 (1.58) 7.3σ

Methanococcus voltae A3 strain A3 5S ribosomal RNA, complete sequence NCBI Reference Sequence: NR_075764.1

TAACGGTCATAGCGGAGGTGTACATCCGATC**CCATTC**CGATCTCGGAAATTAAGCCCTCCAGCGATTTCT
TAA**GTA****CTG****CCATAT****G**GGGAA**CAAGATGA**CGCTGTTAGTCA

13 109 2.34 (1.53) 7σ

Methanocaldococcus jannaschii DSM 2661 strain DSM 2661 5S ribosomal RNA, complete sequence NCBI Reference Sequence: NR_076076.1

TGGTACGGCGGTCATAGCGGGGGGGCCACACCCGAACCCATCCCCGAACCTCGGAAGTTAAGCCCCCAGCG
ATGCCCCGA**GTA****CTG****CCAT**CTGGCGGGAAAGGGGCGACACCGCCGGCCAC

8 116 2.5 (1.58) 3.5σ

Methanosaeta harundinacea 6Ac strain 6Ac 5S ribosomal RNA, complete sequence NCBI Reference Sequence: NR_103229.1 5S

CGAGTGTGGCG**GCCAT**AGCGGCGGGGCAACTCCTGAAC**CCATTC**CGAACTCAG**AAGAT**AAGCCCGCGTGC
GTTCCA**TA****CTG****TACT**TAGGTGCGCGAGCCCTCGGGAACCTATGGATCG**CTGCCA**

8 119 2.55 (1.6) 3.4σ

Methanosarcina mazei 5S ribosomal RNA gene, complete sequence NCBI Reference Sequence: NR_075125.1

AGTTTGGCG**GCCAT**AGCGGCAGGGCAACTCCTGTACCCATCCCGAACACAG**AAGAT**AAGC**CTGCC**CGCGT
TCCT**TA****CTG****TACTGA**AAGTGTGCGAGCCTTCGGGAACCTCTGGATCG**CTGCCA**AGCTCACCT

8 126 2.7 (1.65) 3.2σ

Nanoarchaeum equitans Kin4-M 5S ribosomal RNA, complete sequence NCBI Reference Sequence: NR_103265.1

GGAGG**GCCAT**AGCGGCCCGGGAACCAACCCGTACCCATCTCGAACACGGAAGTTAAGCCGGGCGCGTCCC
GAG**TGGTACTGCC**CCGCGAAGGGGTGGGAAGCTCGGGATGCCCTCCC

7 113 2.4 (1.56) 3σ

Methanosarcina barkeri str. Fusaro strain Fusaro 5S ribosomal RNA, complete sequence NCBI Reference Sequence: NR_075338.1

AGAGTTTGGCG**GCCAT**AGCGGCAGTGTAACCTCCTGTACCCATCCCGAACACAGTAGATAAGC**CTGCC**CGC
GTTCTT**TA****CTG****TACTGA**AAGTGTGCGAGCCTTCGGGAACCTCTGGATCG**CTGCCA**TACTCACC

7 127 2.7 (1.65) 2.6σ

Thermococcus kodakarensis KOD1 strain KOD1 5S ribosomal RNA, complete sequence NCBI Reference Sequence: NR_075222.1

GGTACGGCGGCCATAGCGGCGGGGCCACACCCGGTCTCGTTTCGACCCCGGAAGTTAAGCCCGCCAGCGT
TCCCGGGTGTACTGCCCTCCGAGAGGGGCGGGAAACCGGGAACGCCGCCGGCCAC

6 122 2.6 (1.6) 2.1σ

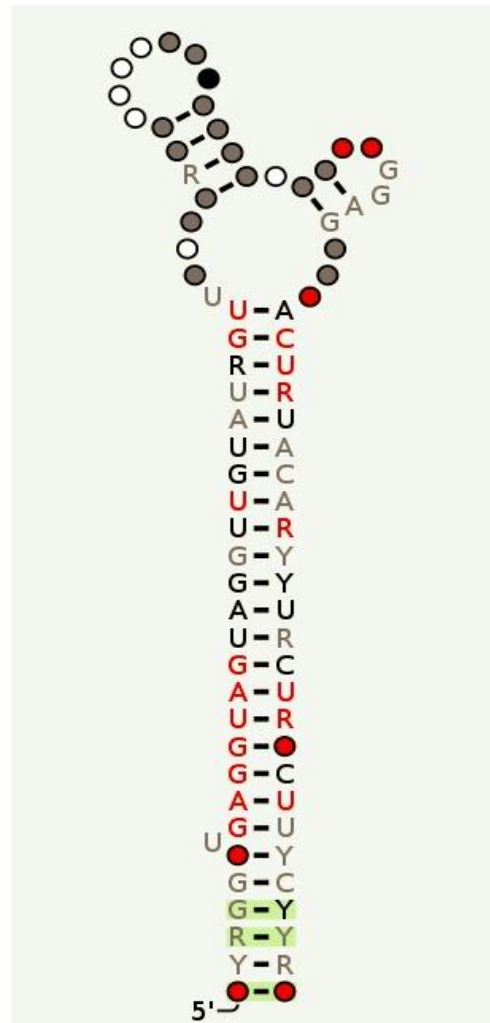


Figure S1. MicroRNA let7 secondary structure