

LA-UR- 09-06514

Approved for public release;
distribution is unlimited.

Title: Biological consequences of ancient gene acquisition and duplication in the large genome of soil bacterium, "Solibacter usitatus" strain Ellin6076

Author(s): Jean F. Challacombe, Stephanie A. Eichorst, Loren Hauser, Miriam Land, Gary Xie, Cheryl Kuske

Intended for: PNAS



Los Alamos National Laboratory, an affirmative action/equal opportunity employer, is operated by the Los Alamos National Security, LLC for the National Nuclear Security Administration of the U.S. Department of Energy under contract DE-AC52-06NA25396. By acceptance of this article, the publisher recognizes that the U.S. Government retains a nonexclusive, royalty-free license to publish or reproduce the published form of this contribution, or to allow others to do so, for U.S. Government purposes. Los Alamos National Laboratory requests that the publisher identify this article as work performed under the auspices of the U.S. Department of Energy. Los Alamos National Laboratory strongly supports academic freedom and a researcher's right to publish; as an institution, however, the Laboratory does not endorse the viewpoint of a publication or guarantee its technical correctness.

**Biological consequences of ancient gene acquisition and duplication in the large
genome of soil bacterium, “*Solibacter usitatus*” strain Ellin6076**

Jean F. Challacombe^{1*}, Stephanie A. Eichorst^{2*}, Loren Hauser³, Miriam Land³, Gary
Xie¹, and Cheryl R. Kuske²

¹Genome Science/DOE Joint Genome Institute, Bioscience Division, Los Alamos
National Laboratory, Los Alamos, NM 87545, ²Bioscience Division, Los Alamos
National Laboratory, Los Alamos, NM 87545, ³Genome Analysis and Systems Modeling
Group, Biosciences Division, Oak Ridge National Laboratory, Oak Ridge, TN 37831.

* Both authors contributed significantly to this study

Corresponding author: Cheryl R. Kuske, Bioscience Division, Los Alamos National
Laboratory, Los Alamos, NM 87545, Email: kuske@lanl.gov, Phone: (505) 665-4800

Classification: BIOLOGICAL SCIENCES: Evolution

Manuscript information: 29 text pages, 1 figure, 3 tables

Data deposition footnote: The genomes described in this manuscript have the GenBank
accession numbers NC_008536 and NC_008009.

ABSTRACT

Bacterial genome sizes range from ca. 0.5 to 10 Mb and are influenced by gene duplication, horizontal gene transfer, gene loss and other evolutionary processes. Sequenced genomes of strains in the phylum *Acidobacteria* revealed that “*Solibacter usistatus*” strain Ellin6076 harbors a 9.9 Mb genome. This large genome appears to have arisen by horizontal gene transfer via ancient bacteriophage and plasmid-mediated transduction, as well as widespread small-scale gene duplications. This has resulted in an increased number of paralogs that are potentially ecologically important (ecoparalogs). Low amino acid sequence identities among functional group members and lack of conserved gene order and orientation in the regions containing similar groups of paralogs suggest that most of the paralogs were not the result of recent duplication events. The genome sizes of cultured subdivision 1 and 3 strains in the phylum *Acidobacteria* were estimated using pulsed-field gel electrophoresis to determine the prevalence of the large genome trait within the phylum. Members of subdivision 1 were estimated to have smaller genome sizes ranging from ca. 2.0 to 4.8 Mb, whereas members of subdivision 3 had slightly larger genomes, from ca. 5.8 to 9.9 Mb. It is hypothesized that the large genome of strain Ellin6076 encodes traits that provide a selective metabolic, defensive and regulatory advantage in the variable soil environment.

INTRODUCTION

Soils contain an abundant and diverse array of bacteria. Based on 16S rRNA gene sequences, members of the phylum *Acidobacteria* are one of the most abundant phyla found in soils and sediments worldwide (1-3). They are also found in a wide variety of other environments, including aquatic (4, 5), extreme (6, 7), polluted (8), and wastewater environments (9, 10). This phylum is defined by a large collection of 16S rRNA gene sequences (>8,000 in the ARB_SILVA Database (July 2009)) (11) that fall into 26 major subdivisions (8). However they have been very difficult to isolate *in vitro*, and few validated cultured representatives exist (7, 12-16). The cultured isolates are slow-growing and can be challenging to grow in the lab, which has hampered biological and physiological characterization. Despite their widespread occurrence in nature, much about acidobacteria biology and potential ecological roles in soil remain unknown.

Comparative analysis of three sequenced acidobacteria genomes from subdivisions 1 and 3 revealed that the genome of the subdivision 3 member, "*Solibacter usistatus*" strain Ellin6076 (hereafter termed Ellin6076), was 9.9 Mb in size whereas subdivision 1 strains, "*Korebacter versitalis*" strain Ellin345 (hereafter termed Ellin345) and *Acidobacterium capsulatum* had much smaller genomes (5.7 Mb and 4.1Mb, respectively) (17). We hypothesized that identifying the mechanisms leading to the large genome of Ellin6076 and distinctive physiological features would provide information about its potential ecological roles in the soil.

Variation in genome size, structure and gene arrangement impacts bacterial phenotype and contributes to genome evolution (18). Genome size can differ dramatically

1 within the same genus or family, and is not associated with specific bacterial lineages or
2 phenotypes. Large genomes (defined here as >7 Mb) are found in many diverse species
3 across the Domain *Bacteria* (17, 19-26). Obligatory host-associated bacterial pathogens
4 (27), insect symbionts (27-29), and extremophiles (30, 31) have lost genes during
5 specialization for their environments (27, 32), and typically harbor reduced genomes
6 compared to facultative pathogens and free-living bacteria (27-31, 33). Free living soil
7 and marine species (26) along with facultative pathogens tend to have larger genomes
8 (34), and are presumed to provide a selective advantage in their environments, which
9 contain scarce, and potentially diverse resources (35). In a highly variable, changing
10 environment it may be advantageous for microorganisms to retain a broader collection of
11 genes in order to respond to environmental perturbations (36).

12 Several mechanisms contribute to genome size, structure and evolution, including
13 whole genome duplication, large tandem duplications, small-scale duplication events, and
14 horizontal gene transfer (18, 27, 36-47). Gene duplication and horizontal gene transfer
15 events are mechanisms for homolog acquisition (36, 37, 39-44, 47), often leading to the
16 generation of paralogs, which can undergo mutations that lead to functional
17 diversification (36, 37, 39, 40, 47). Horizontal transfer events such as natural
18 transformation, conjugative transfer and bacteriophage-mediated transduction can
19 introduce genes from external sources, and may also result in acquisition of new
20 functions (41, 48-50). In this study, the physical features of the large Ellin6076 genome
21 were examined to determine possible mechanisms leading to the generation of the large
22 genome, and to develop hypotheses about the stability, physiology and possible
23 ecological consequences of harboring a large genome. Results from this genomic

approach have identified potential traits encoded in paralogs that suggest physiological drivers for maintaining a large genome in a soil environment.

RESULTS

Mechanisms contributing to genome size and structure

Small scale repeat sequences. The presence of short repeat sequences in the Ellin6076 genome was a key factor contributing to its large size. The total number of repeats in Ellin6076 was 8.7-fold greater than the number of repeats in Ellin345 (supplemental material, Tables S1-S5). The majority of all repeats identified in both genomes were less than 50 - 100 nucleotides long. Consistent with other large bacterial genomes, MUMmer (51) analysis (supplemental material, Figure S1) demonstrated that long repeat regions and/or whole genome duplication did not contribute to the large genome of Ellin6076.

Gene acquisition. Three genomic features were examined that can indicate the presence of horizontally acquired genes associated with genomic islands or prophage regions. These features were deviant GC content compared to the rest of the genome (52), genes encoding mobile genetic elements, and genomic sequence regions with deviant dinucleotide bias, that can indicate the presence of genomic islands (18, 34, 41, 42, 48, 52-57).

GC content. Circular representations of the Ellin6076 and Ellin345 genomes in Figure 1 illustrate the general features of the two genomes, including forward and reverse coding sequences (circles 1 and 2 starting at the outside), RNA genes, locations of mobile elements, GC content, and GC skew. By examining the circular diagram, the cumulative

1 GC skew (58) and output from IslandPath (55), we found that the Ellin6076 genome did
2 not contain any large regions with GC content that differed significantly from the average
3 GC content of the rest of the genome.

4 Mobile genetic elements. The Ellin6076 genome contained 123 mobile genetic
5 element genes that encode phage integrases, transposases and IS elements, compared to
6 29 mobile element genes in Ellin345 (Figure 1). Given the large number of phage
7 integrase genes, bacteriophage-mediated transduction appears to have occurred in
8 Ellin6076, however no intact prophage regions were identified in the genome (17).
9 Numerous integrase, transposase and IS element protein sequences had a high (97–100%)
10 identity to each other, and could be classified into identity groups (supplemental material,
11 Table S6). Given the high sequence similarity of the genes within these mobile element
12 groups, it appears that these genes may have recently duplicated and dispersed
13 throughout the genome.

14 To further investigate the potential for past phage infection, we examined the
15 genome for prophage regions, as well as clustered, regularly interspaced, short
16 palindromic repeats (CRISPRs), which are composed of recurrent direct repeat sequences
17 interspaced by intervening non-repetitive sequences of phage origin, and indicate a
18 previous phage infection event (59-61). We did not find any prophages or CRISPRs in
19 the genome, but identified 3 CRISPR-associated protein genes (Acid_0892, Acid_0893
20 and Acid_0895), which may be involved in DNA repair. These data suggest that the
21 presumed phage transduction events that introduced genes into the Ellin6076 genome did
22 not happen recently, given the lack of identification of any intact prophage regions and
23 CRISPRs, which do not persist in the host genome over time (62).

1 Genomic Islands. A genomic island (GI) is a region of a genome that was
2 previously mobile, and has become fixed in the genome. GIs are frequently inserted near
3 tRNA genes and flanking repeat sequences, and often contain genes encoding mobile
4 elements (reviewed in (53)), as well as niche-specific functions, such as virulence or
5 metabolic traits. Since the Ellin6076 genome contained many genes encoding mobile
6 elements, we examined these regions for candidate GIs. Using the output from IslandPath
7 (55), putative GIs were identified in the Ellin6076 genome (supplemental material, Table
8 S7); GI features identified by IslandPath included gene clusters with dinucleotide bias
9 and GC content that deviated from the average GC content of the genome, and the
10 presence of either tRNA or mobile element genes flanking the putative GI.

11 Putative GI regions in the Ellin6076 genome were difficult to identify since the
12 overall GC content in these regions was not strikingly deviant from the average GC
13 content of the genome. Sixteen candidate GIs were identified, which contained small
14 groups of genes with GC content either above or below the average GC content of the
15 genome along with numerous tRNA genes and mobile genetic elements. The most
16 notable GI candidate regions (III, XI, XIII and XV) (supplemental material, Table S7)
17 were characterized by a below-average GC content, dinucleotide bias that covered nearly
18 the entire region, the presence of tRNA genes, and mobile elements. In addition to the
19 abundant phage integrase genes, some of the putative GI regions contained fragments of
20 genes annotated as either phage or plasmid-related. All of the putative GIs contained
21 multiple genes annotated as hypothetical, with no similarity to any known sequences.

22 Gene duplication and paralogs. Evidence for gene duplication was found in the
23 Ellin6076 genome, and corresponded to an increase in the number of paralogs (duplicated

genes) in virtually all COG categories compared to Ellin345 (Table 1 & supplemental material Tables S8-S9). After correcting for genome size, Ellin6076 contained a higher percentage of paralogs (67.8%, 5426 paralogs in 1103 paralog groups) than Ellin345 (52.6%, 2543 paralogs in 677 paralog groups). The number of paralogs in many categories was increased 4-fold or greater (Tables 2 and S10), and included mobile elements, genes involved in cell wall/membrane biogenesis, signal transduction, intracellular trafficking and secretion, defense mechanisms and metabolism. Paralogous sequences with the same functional definition (e.g. serine/threonine kinases) in Ellin6076, generally showed less than 50% amino acid identity to each other in full-length alignments (data not shown), providing evidence for ancient duplication events.

Examples of increased functional redundancy and diversity in Ellin6076

The functional categories described above represent increased metabolic and regulatory redundancy and diversity encoded in the Ellin6076 genome compared to Ellin345 (Table 2). Increased numbers of paralogs with functions related to multiple aspects of metabolic breadth, cellular defense, and gene regulation were common. Some notable examples of functional categories showing a 4-fold or greater increase in Ellin6076 are described below, and the full list is presented in Table S10.

Metabolic breadth. The metabolic redundancy and diversity present in Ellin6076 included paralogs that were overrepresented in Ellin6076 compared to Ellin345, as well as genes encoding functions that were unique to Ellin6076. Putative functions that were overrepresented in Ellin6076 compared to Ellin345 included numerous metabolic enzymes. Some examples include FAD/FMN-containing dehydrogenases, lipases,

1 choline dehydrogenase, xylose isomerases, gluconolactonase, alpha-L-fucosidase,
2 glucose/sorbose dehydrogenases, along with enzymes utilized in ubiquinone
3 biosynthesis. Examples of increased carbon degradation ability include: 41 genes
4 encoding xylose isomerase (compared to 6 genes in Ellin 345), which performs pentose
5 and glucuronate interconversions, as well as fructose, mannose and glucose metabolism
6 (63); 7 genes for alpha-L-fucosidase, involved in glycoprotein degradation (64); 4 genes
7 encoding beta-galactosidase (beta-galactoside degradation) (65); and 4 genes encoding
8 beta-xylosidase (xylan metabolism) (65);

9 Unlike the genome of Ellin345, Ellin6076 harbored genes involved in aromatic
10 and polymer (ie, carbon) degradation such as 5 genes encoding 2,4-dihydroxyhept-2-ene-
11 1,7-dioic acid aldolase, which is involved in metabolism of aromatic compounds (66) and
12 endopolygalacturonase (11 genes), involved in hydrolysis of 1,4-alpha-D-
13 galactosiduronic linkages in pectate, galacturonans, and degradation of plant cell walls
14 (67). The Ellin6076 genome contained ca. 6.3-fold more genes than Ellin345 in the
15 category of "L-alanine-DL-glutamate epimerase and related enzymes", which appear to
16 encode mandelate racemase/muconate lactonizing enzymes of the enolase superfamily
17 implicated in aromatic acid catabolism (68), as well as all of the genes encoding the
18 protocatechuate catabolism pathway, which were not present in the smaller genome of
19 Ellin345 (supplemental material, Table S12).

20 Other unique metabolic functions encoded in the Ellin6076 genome included: 4
21 genes encoding cytochrome b; quinoprotein glucose dehydrogenase (26 genes), which
22 functions in glucose and glucose-1-phosphate degradation (69); 9 genes annotated as
23 'microcompartments protein', with sequence similarity to ethanolamine and propanediol

1 utilization proteins, which may play a role in modulating toxicity of aldehydes produced
2 during 1,2-propanediol and ethanolamine degradation (70); and other genes involved in
3 ester hydrolysis, sulfur metabolism and menaquinone biosynthesis.

4
5 Cellular defense. Ellin6076 contained genes encoding functions related to
6 antibiotic synthesis and resistance that support an increased potential for cellular defense.
7 In particular, 26 genes encoding 4-amino-4-deoxy-L-arabinose transferase and at least 17
8 genes for glycosyltransferase family 2, which are involved in polymixin antibiotic
9 resistance (71). Both Ellin345 and Ellin6076 encode such glycosyltransferases
10 (Acid345_2286 and Acid_1759) that are orthologs of the *Burkholderia xenovorans*
11 polymixin resistance glycosyltransferase (Bxe_A2387). The Ellin6076 genome also
12 contained 12 genes encoding an antirepressor that regulates drug resistance (peptidase
13 M56), and 4 genes encoding phytanoyl-CoA dioxygenase, which is involved in
14 biosynthesis of mitomycin antibiotics (72). Ellin345 had 1 gene (Acid345_4531)
15 encoding this putative activity, which corresponded to one of the Ellin6076 genes
16 (Acid_4483); orthologs were also found in *B. xenovorans* (Bxe_B0569), *Nocardia*
17 *farcinica* (nfa35020), *Rhodococcus* sp. RHA1 (RHA1_ro05315), and several other soil-
18 dwelling species.

19 Adaptation to environmental conditions. Abundant genes encoding signaling
20 factors and regulators of cellular function are important features that suggest ability of the
21 cell to alter its physiology in response to changing environmental conditions. One such
22 regulator is sigma-24 extracytoplasmic function (ECF), which is also called sigma E (73).
23 ECF sigma-24 factors regulate the expression of proteins in the bacterial outer membrane

(Gram-positive) or periplasmic space (Gram-negative) in response to stressful external stimuli (74-81). The Ellin6076 genome contained 74 sigma-24 ECF factor (sigma E) genes, only 16 of these had counterparts in Ellin345. Thirteen of the Ellin6076 sigma-24 genes were adjacent to an anti-sigma factor gene in an operon-like arrangement, and most of these sigma-anti-sigma pairs were listed as predicted operons on the Microbesonline website (82) (<http://www.microbesonline.org/operons/gnc234267.html>). When present in the same operon, anti-sigma factors are co-transcribed with the ECF sigma factor, inhibiting the activity of the sigma factor and protecting it from degradation so that it is ready to act when needed (83). The co-localization of 13 ECF sigma-24 factor genes with anti-sigma factors in operons indicates that at least these 13 sigma-24 ECF factors are functional in the traditional sense (83). However this does not preclude the possibility that these components and/or the additional ECF sigma-24 factors encoded in the Ellin6076 genome are used in different and novel ways that have not yet been described (74)

Ellin6076 also contained increased numbers of genes encoding two-component systems, serine/threonine protein kinases, transcriptional regulators, and tetratricopeptide/TPR repeat proteins, which may have a variety of functions, including protein-protein interactions and the assembly of multiprotein complexes during a variety of biological processes, such as cell cycle regulation and transcriptional control (84).

Genome size survey of other members from the phylum *Acidobacteria*

Less than six named and described strains exist to represent the phylum *Acidobacteria*, and most soil isolates represent only two of the twenty-six recognized

subdivisions in the phylum (8). Pulsed-field gel electrophoresis was used to estimate the genome size of other acidobacteria strains from subdivisions 1 and 3 to determine if the large genome trait was specific to Ellin6076 or more widely distributed among the cultured isolates. The isolates represented different genera, with 16S rRNA gene sequence similarity (a measure of bacterial relatedness) of ca. 90%, between the subdivision 3 strains evaluated (n=2), and of ca. 92% among the subdivision 1 strains that were tested (n=5; Figure 2). Members of subdivision 1 were found to have smaller size genomes ranging from ca. 2.0 to 4.8 Mb, whereas the two members of subdivision 3 harbored genomes size of ca. 5.8 and 9.9 Mb. (Table 3).

DISCUSSION

Multiple mechanisms were found to contribute to the large genome of Ellin6076, including ancient horizontal acquisition events, probably mediated by phage, and small duplication events. Collectively they resulted in multiple, divergent paralogs that confer the potential for broader and more differentially regulated physiology in Ellin6076. Consistent with studies of other bacterial genomes (85, 86), we did not find evidence for large insertions or large tandem duplications. The potential ecological consequences of the increased capacity for metabolism, defense, and regulation in the Ellin6076 genome are discussed here.

Evidence for ancient horizontal transfer followed by degradation of the signatures of horizontal transfer

1 Our genome analysis results suggest that the extensive gene duplication observed
2 in Ellin6076 may have begun with ancient horizontal acquisition events. The lack of
3 strand bias, increased number of repeats, and distribution pattern of the 126 mobile
4 genetic elements throughout the genome of Ellin6076 suggest that horizontal transfer,
5 followed by gene duplication, repeat-mediated recombination and transposition may have
6 acted to shape the structure of this genome. Horizontal gene transfer is a powerful
7 evolutionary event, evident by the acquisition of new capabilities and thus the possibility
8 to exploit a new niche(s) (87). Natural transformation, conjugative transfer of plasmid
9 sequences, and bacteriophage-mediated transduction can introduce genes into a genome
10 from external sources not only resulting in the acquisition of new function(s) (41, 48-50),
11 but also leading to an increase in genome size.

12 The absence of intact phage or plasmid regions, taken together with the presence
13 of numerous mobile genetic elements (i.e. transposases, integrases, and IS elements),
14 CRISPR-associated protein genes, putative candidate genomic islands with deviant GC
15 content and dinucleotide bias, and phage-related genes (such as phage integrases)
16 illustrate that phage and/or plasmid integration events helped to shape the Ellin6076
17 genome. However, they were not recent and have become obscured by more recent gene
18 duplications and genome rearrangements mediated by the abundant mobile elements and
19 repeats within the genome. Adding to the increasing genomic evidence from other
20 bacterial species, our results support the conjecture that large bacterial genomes result
21 from ancient processes, most likely phage and/or plasmid integration and gene
22 duplication(25). These events generate functional redundancy and diversity within the
23 genome, which persists over time because it offers a survival benefit to the organism.

1 Phages have been observed with a high abundance, ca. 1.9×10^9 virus-like
2 particles/gram soil (dry weight), in forested and agricultural soils (88), and are an
3 important agent of horizontal gene transfer in the soil environment. It is thus reasonable
4 to speculate that ancient soil phage have contributed to the genome diversification of
5 Ellin6076 given the remnants of numerous mobile genetic elements. It does not appear
6 that either Ellin6076 or Ellin345 are currently very susceptible to phage transfection
7 given the lack of evidence for recent phage events.

9 **Ancient gene duplication events resulted in increased number of paralogs**

10 Most of the Ellin6076 paralogs within a particular functional group had relatively
11 low sequence similarities when compared to each other (e.g. serine/threonine protein
12 kinases), suggesting they arose from ancient duplication events and then diverged.
13 Examination of the gene neighborhoods surrounding these divergent paralogs revealed
14 the presence of one or more mobile elements (Table S6), signifying that these paralogous
15 sequences may have arisen through duplication, followed by movement within the
16 genome via nearby mobile elements or through repeat-mediated recombination events.
17 This observation is in agreement with a study of IS elements in cyanobacteria and
18 archaea (89) and genome rearrangements in *Sulfolobus solfataricus* P2 (30) suggesting
19 that IS elements can facilitate genomic changes by transposase-mediated transposition
20 and by increases in copy number through homologous recombination (90) and self-
21 replicating behavior (85). Other studies have documented that the number of recently
22 active IS elements increased with genome size (89), and the regions adjacent to these IS
23 elements were enriched in genes encoding regulatory and metabolic functions. Notably,

1 COG categories representing regulatory and metabolic functions were expanded in the
2 large genome of Ellin6076, and we did find several groups of related mobile element
3 genes with high sequence similarity to each other (Table 6) adjacent to regions containing
4 regulatory and metabolic genes.

5

6 **Abundant paralogs provide functional diversity**

7 Gene duplication in various COG functional categories can result in gene family
8 expansion through functional redundancy with existing genes. In addition, paralogs
9 resulting from gene duplications can also undergo mutations leading to functional
10 diversification (36, 38, 39, 43, 47). This chain of events would lead to expansion of
11 various COG categories, as we observed in the large genome of Ellin6076.

12 The Ellin6076 genome contained an abundance of paralogs that occurred together
13 in more than one area of the genome, but not in the same order and orientation
14 (supplemental material, Table S8), which is consistent with previous studies (47). The
15 genome contained paralogs in virtually all COG functional categories, especially in genes
16 involved in transcription, metabolism, signal transduction and regulation. The adaptive
17 consequences of carrying these abundance paralogs are unknown, but one could reason
18 that a large genome could confer both advantages and disadvantages to bacterial fitness
19 and performance in the environment. Potential advantages include increased breadth and
20 regulatory potential of metabolic function, which would be offset by the metabolic cost to
21 cell growth, in maintenance and replication of the large genome. A large genome that
22 encodes a variety of functionally similar genes that recognize different substrates or that

1 are under different regulatory control could confer substantial metabolic flexibility under
2 variable environmental conditions.

3 The abundant and widely divergent paralogs in the large Ellin6076 genome, found
4 across many aspects of metabolism, defense and regulation, suggest that increased
5 functional diversity is an important criterion in the lifestyle of this species in the soil. To
6 survive in soil, bacteria must cope with extremes of moisture, temperature, and
7 geochemical conditions, and must compete successfully with other microbes for limited,
8 changing resources. Whereas fast growing species may be able to out compete other
9 community members via rapid growth, slow growing organisms like strain Ellin6076
10 must employ other strategies for survival. In culture, Ellin6076 is non-motile, slow-
11 growing (taking weeks to generate visible colonies on culture plates), grows optimally on
12 very low nutrient concentrations (91, 92), and can utilize many different carbon sources
13 (17). This suggests a lifestyle in the soil adapted to relatively stationary, slow growth in
14 nutrient limited conditions. In such a lifestyle, it is consistent to reason that harboring
15 extensively diverse metabolic capability would be advantageous.

16 The increased capacity of Ellin6076 to degrade a variety of carbon suggests that
17 the presence of a large genome is beneficial in soil where carbon sources are highly
18 varied. Notably, the Ellin6076 genome provided increased capability to use complex
19 carbon substrates such as aromatic acids and plant cell wall materials.

20 Specialized microbes that harbor the enzymes to degrade these compounds are essential
21 to the terrestrial carbon cycle since these compounds are slow to degrade.

22 In addition to metabolic exploitation of available resources, Ellin6076's large
23 genome could provide an alternative competitive and survival strategy in adverse

1 environmental conditions, as has been suggested for other soil-dwelling heterotrophs (20,
2 25). For example, the Ellin6076 genome included an expansion of cell wall/membrane
3 biogenesis capabilities and stress response sigma factors, potentially providing an
4 increased ability to protect itself from stressful environmental conditions. The increase in
5 defense mechanism-related genes, most having an assigned function of antimicrobial
6 resistance, could help to defend against neighboring microorganisms that are competing
7 for the same resources.

8 With the increased capacity for carbon metabolism and defense comes an
9 increased requirement for regulation. The Ellin6076 genome harbored increased
10 numbers of genes involved in transcription and signal transduction mechanisms,
11 suggesting an increased ability to regulate metabolism, and/or an increased capability to
12 sense and respond to environmental changes compared to Ellin345 (Table S10). Most
13 notable in these categories were genes encoding serine/threonine protein kinases,
14 transcriptional regulators, and DNA-directed RNA polymerase, sigma-24 (sigma E)
15 homologs. To date, Ellin6076 harbors the most sigma E homologs among the sequenced
16 genomes in the *Bacteria*. Sigma E regulons in a variety of bacteria are induced in
17 response to stressful environmental conditions including nutrient limitation/starvation,
18 oxidative stress, heat shock, lead exposure, cell envelope stress (75-81). Processes
19 activated by SigmaE include outer membrane synthesis and assembly (93), carotenoid
20 biosynthesis (94), mucoid production (95), and organism-specific functions necessary for
21 environmental adaptation (93). It is plausible to assume that Ellin6076's increased
22 capacity to respond to the environment provides a selective advantage in soils in times of
23 stress.

1 The functional redundancy observed in the genome of Ellin6076 might be due to
2 the presence of ecoparalogs (96), which perform the same basic function under different
3 environmental conditions, and can serve to help microorganisms during seasonal periods
4 of fluctuations in resources(97). Patchy nutrient distribution, limiting nutrients, and
5 geochemical conditions that can vary dramatically across the mm scale, are major factors
6 that shape soil microbial communities (98-100). The ability to use varied nutrient sources
7 across gradients of physical conditions would be advantageous in the soil.

8 The functional complexity gained with a large genome may not be unique to
9 microorganisms in the soil environment. An increased bacterial genome size (6 Mb) has
10 also been observed at the border of the oxic and anoxic zones in microbial mats, as
11 compared to the rest of the mat (3 – 3.5 Mb) reflecting the increased functional
12 complexity needed at this depth (101). The large (7.2Mb) genome of the marine
13 bacterium *Hahella chejuensis* also has a number of functionally redundant genes
14 involved in transcriptional regulation and/or environmental sensing that may play roles in
15 its adaptability to a changing marine environment (22).

16 17 **Genome Size Distribution in Related Species**

18 The Ellin6076 has the largest genome of the cultured isolates for which we have
19 genome sizes, and this does not appear to be a trait common to subdivisions 1 and 3. The
20 large genome trait occurs sporadically among closely related species and several large-
21 vs-small genome pairs of related bacteria (usually at the Genus level) have been
22 identified (Table 1). Studies of large genomes greater than 6 Mb in size such as *Hahella*
23 *chejuensis* KCTC 2396 (22), *Bradyrhizobium japonicum* USDA 110 (24),

1 *Mesorhizobium loti* MAFF303099 (23), *Streptomyces coelicolor* A3(2) (19),
2 *Streptomyces avermitilis* MA-4680 (21), *Rhodococcus* sp. RHA1 (25), and *Burkholderia*
3 *xenovorans* LB400 (20) demonstrated increased numbers of duplicate genes among the
4 paralogs, and genome size was strongly correlated with the number of paralogs, which
5 represented particular functional classes of genes involved in adapting to a changing
6 environment (47). The Ellin6076 and Ellin345 pair compared here falls in two different
7 subdivisions within the phylum (Figure 2), and are not as closely related as the other pairs
8 in Table 1. However, these are the two closest large-small genomes available for genome
9 comparisons.

10 The genome analysis presented here has identified features of the large genome
11 that suggest how it was formed and suggests biological consequences that can be inferred
12 from its structure. Clearly, improved culture approaches are needed that will facilitate
13 biochemical and physiological studies suggested by the genome evidence, to ultimately
14 determine the costs and benefits of harboring a large genome.

15

16 MATERIALS AND METHODS

17 Genome sequencing and annotation were described previously (17). We used four
18 freely available programs to assess repeats: repeat-match (finds all repeats), exact-
19 tandems (exact tandem repeats), Tandem Repeats Finder (all tandem repeats, (102)), and
20 Inverted Repeats Finder (inverted repeats, (103)).

21 The MUMmer package (51) programs nucmer, repeat_match and exact_tandems
22 were used for analysis of repeat regions. To identify long inexact genomic repeats, the
23 nucmer program was used with the options –maxmatch and –nosimplify. The resulting

1 dotplot was obtained using mummerplot. The repeat_match and exact Tandems programs
2 were run with default arguments. Tandem repeats finder (102) and inverted repeats
3 finder (103) were run with recommended default arguments to identify tandem and
4 inverted repeats, respectively.

5 COG (104) category distributions, paralog and ortholog groups were obtained
6 from the Integrated Microbial Genomes (IMG) system (<http://img.doe.gov>). In IMG,
7 genes were assigned to Clusters of Orthologous Groups of proteins (COGs) using
8 Reverse Position Specific BLAST of the Ellin6076 sequences against NCBI's Conserved
9 Domain Database as described in (<http://img.jgi.doe.gov/w/doc/userGuide.pdf>).
10 Paralogous pairwise relationships were computed as reciprocal hits within the genome,
11 and paralogous groups were identified using the Markov Cluster Algorithm (MCL) with
12 default parameters (see <http://img.jgi.doe.gov/w/doc/userGuide.pdf>). Comparative
13 analyses were performed using IMG (105, 106), BLAST (107), the Conserved Domain
14 Database (104), MetaCyc (<http://metacyc.org>) and the Pathway Tools (108). Putative
15 genomic islands were identified using IslandPath (55). IMG (105, 106) was used to
16 search for clustered regularly interspaced short palindromic repeats (CRISPRs).

17 The genome size strains from subdivisions 1 and 3 of the phylum *Acidobacteria*
18 was determined using pulsed-field gel electrophoresis (PFGE) upon the restriction of the
19 genome DNA with rare-cutting enzymes (SmaI and PmeI) (New England Biolabs,
20 Beverly, MA) following the procedure described by Bergthorsson and Ochman (109)
21 Restricted genomic DNA was separated by electrophoresis on a CHEF-DR apparatus
22 (Bio-Rad Laboratories, Richmond, CA) including markers such as Yeast Chromosome,
23 Lambda Ladder, and Low Range (New England Biolabs, Beverly, MA).

REFERENCES

1. Barns SM, Takala SL, & Kuske CR (1999) Wide Distribution and Diversity of Members of the Bacterial Kingdom *Acidobacterium* in the Environment. *Appl Environ Microbiol* 65:1731-1737.
2. Dunbar J, Takala S, Barns SM, Davis JA, & Kuske CR (1999) Levels of bacterial community diversity in four arid soils compared by cultivation and 16S rRNA gene cloning. *Appl Environ Microbiol* 65:1662-1669.
3. Stott MB, *et al.* (2008) Isolation of novel bacteria, including a candidate division, from geothermal soil in New Zealand. *Environ Microbiol* 10:2030-2041.
4. Pham VD, Konstantinidis KT, Palden T, & DeLong EF (2008) Phylogenetic analyses of ribosomal DNA-containing bacterioplankton genome fragments from a 4000 m vertical profile in the North Pacific Subtropical Gyre. *Environ Microbiol* 10:2313-2330.
5. Quaiser A, *et al.* (2008) Comparative analysis of genome fragments of *Acidobacteria* from deep Mediterranean plankton. *Environ Microbiol* 10:2704-2717.
6. Hobel CFV, Marteinsson VT, Hreggvidsson GO, & Kristjansson JK (2005) Investigation of the microbial ecology of intertidal hot springs by using diversity analysis of 16S rRNA and chitinase genes. *Appl Environ Microbiol* 71:2771-2776.
7. Kishimoto N, Kosako Y, & Tano T (1991) *Acidobacterium capsulatum* gen. nov., sp. nov.: an acidophilic chemoorganotrophic bacterium containing menaquinone from acidic mineral environment. *Curr Microbiol* 22:1-7.
8. Barns SM, Cain EC, Sommerville L, & Kuske CR (2007) Acidobacteria Phylum Sequences in Uranium-Contaminated Subsurface Sediments Greatly Expand the Known Diversity within the Phylum. *Appl Environ Microbiol* 73:3113-3116.
9. Crocetti GR, Banfield JF, Keller J, Bond PL, & Blackall LL (2002) Glycogen-accumulating organisms in laboratory-scale and full-scale wastewater treatment processes. *Microbiology* 148:3353-3364.
10. LaPara TM, Nakatsu CH, Pantea L, & Alleman JE (2000) Phylogenetic analysis of bacterial communities in mesophilic and thermophilic bioreactors treating pharmaceutical wastewater. *Appl Environ Microbiol* 66:3951-3959.
11. Pruesse E, *et al.* (2007) SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. *Nucleic Acids Res.* 35:7188-7196.
12. Eichorst SA, Breznak JA, & Schmidt TM (2007) Isolation and characterization of newly isolated soil bacteria that define *Terriglobus* gen. nov., in the phylum *Acidobacteria*. *Appl Environ Microbiol* 73:2708-2717.
13. Koch IH, Gich F, Dunfield PF, & Overmann J (2008) *Edaphobacter modestus* gen. nov., sp. nov. and *Edaphobacter aggregans* sp. nov., two novel acidobacteria isolated from alpine and forest soils. *Int J Syst Bacteriol* 58:1114-1122.

- 1 14. Fukunaga Y, Kurahashi M, Yanagi K, Yokota A, & Harayama S (2008)
2 *Acanthopleuribacter pedis* gen. nov., sp. nov., a marine bacterium isolated from a
3 chiton, and description of *Acanthopleuribacteraceae* fam. nov.,
4 *Acanthopleuribacteriales* ord. nov., *Holophagaceae* fam. nov., *Holophagales* ord.
5 nov. and *Holophagae* classis nov. in the phylum 'Acidobacteria'. *Int J Syst Evol*
6 *Microbiol* 58:2597-2601.
- 7 15. Coates J, Ellis DJ, Gaw CV, & Lovley DR (1999) *Geothrix fermentans* gen. nov.,
8 sp. nov., a novel Fe (III)-reducing bacterium from a hydrocarbon-contaminated
9 aquifer. *Int J Syst and Evol Microbiol* 49:1615-1622.
- 10 16. Liesack W, Bak F, Kreft J-U, & Stackebrandt E (1994) *Holophaga foetida* gen.
11 nov., sp. nov., a new, homoacetogenic bacterium degrading methoxylated
12 aromatic compounds. *Arch Microbiol* 162:85-90.
- 13 17. Ward NL, *et al.* (2009) Three Acidobacteria genomes provide a first glimpse of
14 their lifestyles in soils and sediments. *Appl Environ Microbiol* 75:2046-2056.
- 15 18. Darling AE, Miklos I, & Ragan MA (2008) Dynamics of Genome Rearrangement
16 in Bacterial Populations. *PLoS Genet* 4:e1000128.
- 17 19. Bentley SD, *et al.* (2002) Complete genome sequence of the model actinomycete
18 *Streptomyces coelicolor* A3(2). *Nature* 417:141-147.
- 19 20. Chain PSG, *et al.* (2006) *Burkholderia xenovorans* LB400 harbors a multi-
20 replicon, 9.73-Mbp genome shaped for versatility. *Proc Natl Acad Sci USA*
21 103:15280-15287.
- 22 21. Ikeda H, *et al.* (2003) Complete genome sequence and comparative analysis of the
23 industrial microorganism *Streptomyces avermitilis*. *Nat Biotechnol* 21:526-531.
- 24 22. Jeong H, *et al.* (2005) Genomic blueprint of *Hahella chejuensis*, a marine microbe
25 producing an algicidal agent. *Nucleic Acids Res.* 33:7066-7073.
- 26 23. Kaneko T, *et al.* (2000) Complete genome structure of the nitrogen-fixing
27 symbiotic bacterium *Mesorhizobium loti*. *DNA Res.* 7:331-338.
- 28 24. Kaneko T, *et al.* (2002) Complete genomic sequence of nitrogen-fixing symbiotic
29 bacterium *Bradyrhizobium japonicum* USDA110. *DNA Res.* 9:189-197.
- 30 25. McLeod MP, *et al.* (2006) The complete genome of *Rhodococcus* sp. RHA1
31 provides insights into a catabolic powerhouse. *Proc Natl Acad Sci USA*
32 103:15582-15587.
- 33 26. Wackett LP (2008) Large bacterial genomes, > 9 Mb. *Environ Microbiol*
34 10:3423-3424.
- 35 27. Moran NA (2002) Microbial minimism: Genome Reduction in Bacterial
36 Pathogens. *Cell* 108:583-586.
- 37 28. Canback B, Tamas I, & Andersson SGE (2004) A phylogenomic study of
38 endosymbiotic bacteria. *Mol Biol Evol* 21:1110-1122.
- 39 29. Wu D, *et al.* (2006) Metabolic Complementarity and genomics of the Dual
40 bacterial Symbiosis of Sharpshooters. *PLoS Biol.* 4:1079-1092.
- 41 30. She Q, *et al.* (2001) The complete genome of the crenarchaeon *Sulfolobus*
42 *solfataricus* P2. *Proc Natl Acad Sci USA* 98:7835-7840.
- 43 31. Valdés J, Veloso F, Jedlicki E, & Holmes D (2003) Metabolic reconstruction of
44 sulfur assimilation in the extremophile *Acidithiobacillus ferrooxidans* based on
45 genome analysis. *BMC Genomics* 4:51-66.

- 1 32. Ochman H & Moran NA (2001) Genes lost and genes found: evolution of
2 bacterial pathogenesis and symbiosis. *Science* 292:1096-1099.
- 3 33. Baumann C, Judex M, Huber H, & Wirth R (1998) Estimation of genome size of
4 hyperthermophiles. *Extremophiles* 2:101-108.
- 5 34. Ochman H & Davalos LM (2006) The nature and dynamics of bacterial genomes.
6 *Science* 311:1730-1733.
- 7 35. Konstantinidis KT & Tiedje JM (2004) Trends between gene content and genome
8 size in prokaryotic species with larger genomes. *Proc Natl Acad Sci USA*
9 101:3160-3165.
- 10 36. Ranea JAG, Buchan DWA, Thornton JM, & Orengo CA (2003) Evolution of
11 protein Superfamilies and bacterial Genome Size. *J Mol Biol* 336:871-887.
- 12 37. Chothia C, Gough J, Vogel C, & Teichmann SA (2003) Evolution of the protein
13 repertoire. *Science* 300:1701-1703.
- 14 38. Koonin EV, Wolf YI, & Karev GP (2002) The structure of the protein universe
15 and genome evolution. *Nature Rev Microbiol* 420:218-223.
- 16 39. Jordan IK, Makarova KS, Spouge JL, Wolf YI, & Koonin EV (2001) Lineage-
17 specific gene expansions in bacterial and archaeal genomes. *Genome Res.* 11:555-
18 565.
- 19 40. Lynch M & Conery JS (2000) The evolutionary fate and consequences of
20 duplicate genes. *Science* 290:1151-1155.
- 21 41. Frost LS, Leplae R, Summers AO, & Toussaint A (2005) Mobile genetic
22 elements: Agents of open source evolution. *Nature Rev Microbiol* 3:722-732.
- 23 42. Lawrence JG & Ochman H (2002) Reconciling the many faces of lateral gene
24 transfer. *Trends Microbiol.* 10:1-4.
- 25 43. Hooper SD & Berg OG (2003) On the nature of gene innovation: Duplication
26 patterns in microbial genomes. *Mol Biol Evol* 20:945-954.
- 27 44. Hooper SD & Berg OG (2003) Duplication is more common among laterally
28 transferred genes than among indigenous genes. *Genome Biol.* 4:R48-56.
- 29 45. Mirkin BG, Fenner TI, Galperin MY, & Koonin EV (2003) Algorithms for
30 computing parsimonious evolutionary scenarios for genome evolution, the last
31 universal common ancestor and dominance of horizontal gene transfer in the
32 evolution of prokaryotes. *BMC Evol Biol* 3:2.
- 33 46. Sugaya N, *et al.* (2004) Causes for the large genome size in a cyanobacterium
34 *Anabaena* sp. PCC7120. *Genome Inform* 15:229-238.
- 35 47. Gevers D, Vandepoele K, Simillion C, & Van de Peer Y (2004) Gene duplication
36 and biased functional retention of paralogs in bacterial genomes. *Trends*
37 *Microbiol.* 12:148-154.
- 38 48. Gogarten JP & Townsend JP (2005) Horizontal gene transfer, genome innovation
39 and evolution. *Nat Rev Microbiol* 3:679-687.
- 40 49. Zaneveld JR, Nemergut DR, & Knight R (2008) Are all horizontal gene transfers
41 created equal? Prospects for mechanism-based studies of HGT patterns.
42 *Microbiology* 154:1-15.
- 43 50. Thomas CM & Nielsen KM (2005) Mechanisms of, and barriers to, horizontal
44 gene transfer between bacteria. *Nat Rev Microbiol* 3:711-721.
- 45 51. Kurtz S, *et al.* (2004) Versatile and open software for comparing large genomes.
46 *Genome Biol.* 5:R12.

- 1 52. Lawrence JG & Ochman H (1997) Amelioration of bacterial genomes: rates of
2 change and exchange. *J Mol Evol* 44:383-397.
- 3 53. Dobrindt U, Hochhut B, Hentschel U, & Hacker J (2004) Genomic islands in
4 pathogenic and environmental microorganisms. *Nat Rev Microbiol* 2:414-424.
- 5 54. Eisen JA (2000) Horizontal gene transfer among microbial genomes: new insights
6 from complete genome analysis. *Curr Opin Genet Dev.* 10:606-611.
- 7 55. Hsiao W, Wan I, Jones SM, & Brinkman FS (2003) IslandPath: aiding detection
8 of genomic islands in prokaryotes. *Bioinformatics* 19:418-420.
- 9 56. Rocha EPC (2004) The replication-related organization of bacterial genomes.
10 *Microbiology* 150:1609-1627.
- 11 57. Karlin S (2001) Detecting anomalous gene clusters and pathogenicity islands in
12 diverse bacterial genomes. *Trends Microbiol.* 9:335-343.
- 13 58. Grigoriev A (1998) Analysing genomes with cumulative skew diagrams. *Nucleic*
14 *Acids Res.* 26:2286-2290.
- 15 59. Mojica FJM, Diez-Villasenor C, Garcia-Martinez J, & Soria E (2005) Intervening
16 sequences of regularly spaced prokaryotic repeats derive from foreign genetic
17 elements. *J Mol Evol.* 60:174-182.
- 18 60. Jansen R, van Embden JDA, Gaastra W, & Schouls LM (2002) Identification of
19 genes that are associated with DNA repeats in prokaryotes. *Mol Microbiol*
20 43:1565-1575.
- 21 61. Sorek R, Kunin V, & Hugenholtz P (2008) CRISPR--a widespread system that
22 provides acquired resistance against phages in bacteria and archaea. *Nat Rev*
23 *Microbiol* 6:181-186.
- 24 62. Haft DH, Selengut JD, Mongodin EF, & Nelson KE (2005) A guild of 45
25 CRISPR-associated (Cas) protein families and multiple CRISPR/Cas subtypes
26 exist in prokaryotic genomes. *PLoS Comput Biol* 1:474-483.
- 27 63. Häusler H, Weber H, & Stütz AE (2001) D-xylose (D-glucose) isomerase (EC
28 5.3.1.5): observations and comments concerning structural requirements of
29 substrates as well as mechanistic features. *J Carbohydr Chem* 20:239-256.
- 30 64. Katayama T, Fujita K, & Yamamoto K (2005) Novel bifidobacterial glycosidases
31 acting on sugar chains of mucin glycoproteins. *J Biosci Bioeng.* 99:457-465.
- 32 65. Hansen W, Schoutens E, & Yourassowsky E (1979) Significance of the detection
33 of beta-galactosidase and of beta-xylosidase in the taxonomic study of the genus
34 *Haemophilus*. *J Clin Pathol.* 32:1058-1060.
- 35 66. Stringfellow JM, Turpin B, & Cooper RA (1995) Sequence of the *Escherichia*
36 *coli* C homoprotocatechuic acid degradative operon completed with that of the
37 2,4-dihydroxyhept-2-ene-1,7-dioic acid aldolase-encoding gene (hpcH). *Gene*
38 166:73-76.
- 39 67. Federici L, *et al.* (2001) Structural requirements of endopolygalacturonase for the
40 interaction with PGIP (polygalacturonase-inhibiting protein). *Proc Natl Acad Sci*
41 *USA* 98:13425-13430.
- 42 68. Babbitt PC, *et al.* (1996) The enolase superfamily: a general strategy for enzyme-
43 catalyzed abstraction of the alpha-protons of carboxylic acids. *Biochemistry*
44 35:16489-16501.

69. Matsushita K & Ameyama M (1982) D-Glucose dehydrogenase from *Pseudomonas fluorescens*, membrane-bound. *Methods Enzymol.* 89(Part D):149-154.
70. Bobik TA, Havemann GD, Busch RJ, Williams DS, & Aldrich HC (1999) The propanediol utilization (pdu) operon of *Salmonella enterica* serovar Typhimurium LT2 includes genes necessary for formation of polyhedral organelles involved in coenzyme B(12)-dependent 1, 2-propanediol degradation. *J Bacteriol* 181:5967-5975.
71. Trent MS, Ribeiro AA, Lin S, Cotter RJ, & Raetz CRH (2001) An inner membrane enzyme in *Salmonella* and *Escherichia coli* that transfers 4-amino-4-deoxy-L-arabinose to lipid A: induction on polymyxin-resistant mutants and role of a novel lipid-linked donor. *J Biol Chem* 276:43122-43131.
72. García-Estépa R, *et al.* (2006) The *ectD* gene, which is involved in the synthesis of the compatible solute hydroxyectoine, is essential for thermoprotection of the halophilic bacterium *Chromohalobacter salexigens*. *J Bacteriol* 188:3774-3784.
73. Maeda H, Jishage M, Nomura T, Fujita N, & Ishihama A (2000) Two extracytoplasmic function sigma subunits, sigma(E) and sigma(FecI), of *Escherichia coli*: promoter selectivity and intracellular levels. *J Bacteriol* 182:1181-1184.
74. Brooks BE & Buchanan SK (2008) Signaling mechanisms for activation of extracytoplasmic function (ECF) sigma factors. *Biochim Biophys Acta* 1778:1930-1945.
75. Devescovi G & Venturi V (2006) The *Burkholderia cepacia* *rpoE* gene is not involved in exopolysaccharide production and onion pathogenicity. *Can J Microbiol* 52:260-265.
76. Kenyon WJ, Thomas SM, Johnson E, Pallen MJ, & Spector MP (2005) Shifts from glucose to certain secondary carbon-sources result in activation of the extracytoplasmic function sigma factor sigmaE in *Salmonella enterica* serovar Typhimurium. *Microbiology* 151:2373-2383.
77. Manganelli R, Voskuil MI, Schoolnik GK, & Smith I (2001) The *Mycobacterium tuberculosis* ECF sigma factor sigmaE: role in global gene expression and survival in macrophages. *Mol Microbiol* 41:423-437.
78. Potvin E, Sanschagrin F, & Levesque RC (2008) Sigma factors in *Pseudomonas aeruginosa*. *FEMS Microbiol Rev* 32:38-55.
79. Raivio TL & Silhavy TJ (2001) Periplasmic stress and ECF sigma factors. *Annu Rev Microbiol.* 55:591-624.
80. Taghavi S, *et al.* (2009) Lead(II) resistance in *Cupriavidus metallidurans* CH34: interplay between plasmid and chromosomally-located functions. *Antonie Van Leeuwenhoek* 96:171-182.
81. Testerman TL, *et al.* (2002) The alternative sigma factor sigmaE controls antioxidant defences required for *Salmonella* virulence and stationary-phase survival. *Mol Microbiol* 43:771-782.
82. Alm EJ, *et al.* (2005) The MicrobesOnline Web site for comparative genomics. *Genome Res.* 15:1015-1022.
83. Helmann JD (2002) The extracytoplasmic function (ECF) sigma factors. *Adv Microb Physiol.* 46:47-100.

84. Blatch GL & Lässle M (1999) The tetratricopeptide repeat: a structural motif mediating protein-protein interactions. *Bioessays* 21:932-939.
85. Achaz G, Coissac E, P. N, & Rocha EPC (2003) Associations Between Inverted Repeats and the Structural Evolution of Bacterial Genomes. *Genetics* 164:1279-1289.
86. Ochman H, Daubin V, & Lerat E (2004) A bunch of fun guys: the whole-genome view of yeast evolution. *Trends Genet.* 21:1-3.
87. Lawrence JG (1999) Gene transfer, speciation, and the evolution of bacterial genomes. *Curr Opin Microbiol* 2:519-523.
88. Williamson KE, Radosevich M, & Wommack KE (2005) Abundance and diversity of viruses in six Delaware soils. *Appl Environ Microbiol.* 71(6):3119-3125.
89. Zhou F, Olman V, & Xu Y (2008) Insertion Sequences show diverse and recent activities in Cyanobacteria and Archaea. *BMC Genomics* 9:36-46.
90. Redder P & Garrett RA (2006) Mutations and rearrangements in the genome of *Sulfolobus solfataricus* P2. *J Bacteriol* 188:4198-4206.
91. Davis KE, Joseph SJ, & Janssen PH (2005) Effects of growth medium, inoculum size, and incubation time on culturability and isolation of soil bacteria. *Appl Environ Microbiol* 71:826-834.
92. Janssen PH, Yates PS, Grinton BE, Taylor PM, & Sait M (2002) Improved culturability of soil bacteria and isolation in pure culture of novel members of the divisions *Acidobacteria*, *Actinobacteria*, *Proteobacteria*, and *Verrucomicrobia*. *Appl Environ Microbiol* 68:2391-2396.
93. Rhodius VA, Suh WC, Nonaka G, West J, & Gross CA (2006) Conserved and variable functions of the sigmaE stress response in related genomes. *PLoS Biol.* 4:e2.
94. Thirunavukkarasu N, *et al.* (2008) An extra-cytoplasmic function sigma factor and anti-sigma factor control carotenoid biosynthesis in *Azospirillum brasilense*. *Microbiology* 154:2096-2105.
95. Martin DW, Holloway BW, & Deretic V (1993) Characterization of a locus determining the mucoid status of *Pseudomonas aeruginosa*: AlgU shows sequence similarities with a *Bacillus* sigma factor. *J Bacteriol* 175:1153-1164.
96. Sanchez-Perez G, Mira A, Nyiro G, Pasić L, & Rodriguez-Valera F (2008) Adapting to environmental changes using specialized paralogs. *Trends Genet.* 24:154-158.
97. van Elsas JD, Trevors JT, & Wellington EMH (1997) *Modern Soil Microbiology* (Marcel Dekker Inc., New York).
98. Hansel CM, Fendorf S, Jardine PM, & Francis CA (2008) Changes in bacterial and archaeal community structure and functional diversity along a geochemically variable soil profile. *Appl Environ Microbiol.* 74:1620-1633.
99. Zhou J, *et al.* (2002) Spatial and resource factors influencing high microbial diversity in soil. *Appl Environ Microbiol* 68:326-334.
100. Torsvik V & Ovreas L (2002) Microbial diversity and function in soil: from genes to ecosystems. *Curr Opin Microbiol* 5:240-245.
101. Kunin V, *et al.* (2008) Millimeter-scale genetic gradients and community-level molecular convergence in a hypersaline microbial mat. *Mol Syst Biol* 4:198.

- 1 102. Benson G (1999) Tandem repeats finder: a program to analyze DNA sequences.
2 *Nucleic Acids Res.* 27:573-580.
- 3 103. Warburton PE, Giordano J, Cheung F, Gelfand Y, & Benson G (2004) Inverted
4 Repeat Structure of the Human Genome: The X-Chromosome Contains a
5 Preponderance of Large, Highly Homologous Inverted Repeats That Contain
6 Testes Genes. *Genome Res.* 14:1861-1869.
- 7 104. Marchler-Bauer A, *et al.* (2002) CDD: A Database of Conserved Domain
8 Alignments with Links to Domain Three-Dimensional Structure. *Nucleic Acids*
9 *Res* 30:281-283.
- 10 105. Markowitz VM, *et al.* (2006) The Integrated Microbial Genomes (IMG) system.
11 *Nucleic Acids Res* 34:D344-D348.
- 12 106. Markowitz VM, *et al.* (2008) The Integrated Microbial Genomes (IMG) system in
13 2007: data content and analysis tool extensions. *Nucleic Acids Res* 36:D528-
14 D533.
- 15 107. Altschul SF, Gish W, Miller W, Myers EW, & Lipman DJ (1990) Basic local
16 alignment search tool. *J Mol Biol* 215:403-410.
- 17 108. Karp P, Paley S, & Romero P (2002) The Pathway Tools Software.
18 *Bioinformatics* 18:S225-S232.
- 19 109. Bergthorsson U & Ochman H (1995) Heterogeneity of genome sizes among
20 natural isolates of *Escherichia coli*. *J Bacteriol* 177:5784-5789.
21
22
23

FIGURE LEGENDS

Figure 1. Circular map of the Ellin6076 (Panel A) and Ellin345 (Panel B) genomes, showing the distribution of coding sequences, RNA genes, mobile elements, GC content and GC skew. From outside to the center: Circles 1 and 2: forward and reverse strand genes colored by COG categories; Circles 3 and 4: RNA genes (tRNAs green, sRNAs red, other RNAs black); Circles 5 and 6: mobile elements; Circle 7: GC content; Circle 8: GC skew.

COG Code	COG Function Definition
[A]	RNA processing and modification
[B]	Chromatin structure and dynamics
[C]	Energy production and conversion
[D]	Cell cycle control, cell division, chromosome partitioning
[E]	Amino acid transport and metabolism
[F]	Nucleotide transport and metabolism
[G]	Carbohydrate transport and metabolism
[H]	Coenzyme transport and metabolism
[I]	Lipid transport and metabolism
[J]	Translation, ribosomal structure and biogenesis
[K]	Transcription
[L]	Replication, recombination and repair
[M]	Cell wall/membrane/envelope biogenesis
[N]	Cell motility
[O]	Posttranslational modification, protein turnover, chaperones
[P]	Inorganic ion transport and metabolism
[Q]	Secondary metabolites biosynthesis, transport and catabolism
[R]	General function prediction only
[S]	Function unknown
[T]	Signal transduction mechanisms
[U]	Intracellular trafficking, secretion, and vesicular transport
[V]	Defense mechanisms
[W]	Extracellular structures
[Y]	Nuclear structure
[Z]	Cytoskeleton

Figure 2. Maximum-likelihood tree of the phylum *Acidobacteria* subdivisions 1, 2, 3, 4, 5, and 6 (indicated to the right of the group) based on the 16S rRNA gene using

1 sequences obtained from cultivated representatives and environmental clones.
2 Subdivisions 7 and 8 were used as an outgroup (not shown). Strains in which the genome
3 size was determined are highlighted in bold typeface. Each gray trapezoid represents a
4 different subdivision with approximately 5 representative sequences. Internal nodes
5 support by a bootstrap value of >95% are indicated with a filled circle and > 75% with an
6 open circle. The scale bar indicates 0.05 changes per nucleotide.
7

Table 1. Paralogs in large-small genome pairs#

Classification	Large genome	Size (Mb)	# genes	# paralogs (% of genes)	Small genome	Size (Mb)	# genes	# paralogs (% of genes)
Acidobacteria	Strain Ellin 6076	9.97	8002	5426 (67.8%)	Strain Ellin 345	5.65	4834	2543 (52.6%)
Alpha proteobacteria	<i>Magnetospirillum magnetotacticum</i> MS-1*	9.21	10334	6113 (59.2%)	<i>Magnetospirillum magneticum</i> AMB-1	4.97	4667	2428 (52.0%)
Alpha proteobacteria	<i>Mesorhizobium loti</i> MAFF303099	7.6	7352	4396 (59.8%)	<i>Mesorhizobium</i> sp. BNC1	4.94	4686	2606 (55.6%)
Beta proteobacteria	<i>Ralstonia eutropha</i> H16	7.42	6702	4395 (65.6%)	<i>Ralstonia solanacearum</i> UW551*	5.9	4418	2267 (51.3%)
Gamma proteobacteria	<i>Hahella chejuensis</i> KCTC 2396	7.22	6862	3242 (47.2%)	<i>Oceanospirillum</i> sp. MED92*	3.87	3735	1721 (46.1%)
Actinobacteria	<i>Mycobacterium smegmatis</i> MC2 155	6.99	6925	4535 (65.5%)	<i>Mycobacterium leprae</i> TN	3.27	2752	490 (17.8%)

#Data from the Integrated Microbial Genomes (IMG) System; <http://img.jgi.doe.gov>

*Draft genome data

Table 2. Distribution of genes in COG categories for *Acidobacteria* strains Ellin 6076 and Ellin 345. Only categories with four-fold or greater differences are shown.

COG CATEGORY	Strain Ellin 6076	Strain Ellin 345	Fold Increase+
Information storage and processing			
(L) Replication, recombination, repair			
Site-specific recombinase XerD	32	6	5.3
Transposase and inactivated derivatives	64	10	6.4
Cellular processes			
(M) Cell wall/membrane biogenesis			
4-amino-4-deoxy-L-arabinose transferase and related	26	7	3.7
L-alanine-DL-glutamate epimerase and related	25	4	6.3
Periplasmic protease	13	2	6.5
Endopolygalacturonase	11	0	> 11.0
Dihydropicolinate synthase/N-acetylneuraminate lyase	9	2	4.5
ABC-type polysaccharide/polyol phosphate export system, permease	6	0	> 6.0
ABC-type polysaccharide/polyol phosphate export system, ATPase	6	0	> 6.0
Membrane proteins related to metalloendopeptidases	5	1	5.0
Sortase	6	1	6.0
(T) Signal transduction mechanisms			
Antirepressor regulating drug resistance, signal transduction comp.	12	3	4.0
Bacteriophytochrome	9	2	4.5
(U) Intracellular trafficking, secretion			
Flp pilus assembly protein, ATPase CpaE	4	0	> 4.0
Flp pilus assembly protein, ATPase CpaF	4	0	> 4.0
(C) Energy production, conversion			
FAD/FMN-containing dehydrogenases	8	2	4.0
Carbon dioxide conc. mechanism/carboxysome shell proteins	9	0	> 9.0
FOG:HEAT repeat	6	0	> 6.0
Rieske Fe-S protein	5	0	> 5.0
Predicted acetamidase/formamidase	4	0	> 4.0
Cytochrome b subunit	4	0	> 4.0
(E) Amino acid transport, metabolism			
Lysophospholipase L1 and related esterases	16	1	16.0
Dihydropicolinate synthase/N-acetylneuraminate lyase	9	2	4.5
Choline dehydrogenase and related	8	2	4.0
Spermidine synthase	5	0	> 5.0
Asparagine synthase (glutamate hydrolyzing)	4	0	> 4.0
(G) Carbohydrate transport, metabolism			
Sugar phosphate isomerases/epimerases	41	6	6.8
Glucose dehydrogenase	26	0	> 26.0
Gluconolactonase	13	2	6.5
Alpha-L-arabinofuranosidase	8	2	4.0
Alpha-L-fucosidase	7	1	7.0
Glucose/sorbose dehydrogenases	6	1	6.0
ABC-type polysaccharide/polyol phosphate export system, permease	6	0	> 6.0
ABC-type polysaccharide/polyol phosphate export system, ATPase	6	0	> 6.0
2,4-dihydroxyhept-2-ene-1,7-dioic acid aldolase	5	0	> 5.0
Beta-xylosidase	4	0	> 4.0
Beta-galactosidase	4	1	4.0
(H) Coenzyme transport, metabolism			
2-polyprenyl-3-methyl-5-hydroxy-6-methoxy-1,4-benzoquinol methylase	18	4	4.5
Demethylmenaquinone methyltransferase	7	0	> 7.0
(I) Lipid transport, metabolism			
Carboxylesterase type B	9	0	> 9.0
(P) Inorganic ion transport, metabolism			
Arylsulfatase A and related enzymes	18	0	> 18.0
Enterochelin esterase and related enzymes	16	2	8.0
Cytochrome c peroxidase	6	0	> 6.0
(Q) Secondary metabolites			
Dienelactone hydrolase and related enzymes	8	1	8.0
Carbon dioxide conc. mechanism/carboxysome shell proteins	9	0	> 9.0
Protein involved in biosynthesis of mitomycin antibiotics/fumonisin	4	0	> 4.0
Predicted enzyme involved in methoxymalonyl-ACP biosynthesis	4	1	4.0

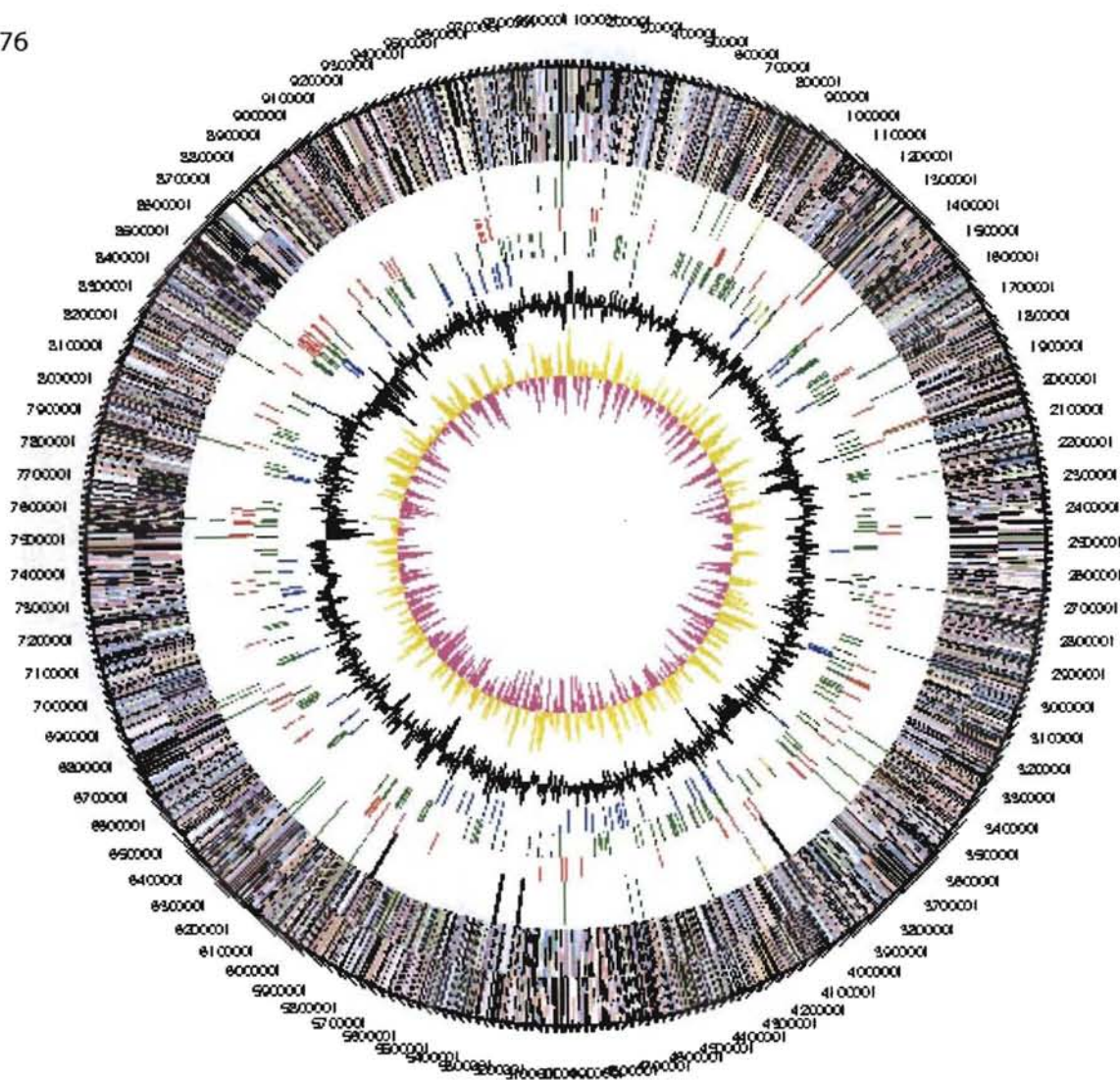
+ Not normalized for genome size

Table 3. Average genome size of representative acidobacterial strains from subdivisions 1, 3 and 4 determined by pulse field gel electrophoresis.

Strain	Mean genomic size $\pm$ std. dev. (Kb) (restriction enzyme used, number of independent determinations)	Strain reference
Subdivision 1		
<i>A. capsulatum</i>	4,128	Ward et al., 2009
" <i>K. versatilis</i> ", strain Ellin345	5,650	Ward et al., 2009
Strain TAA 166	4,036 $\pm$ 325 (<i>SwaI</i> , n=2) 3,817 $\pm$ 128 (<i>PmeI</i> , n=2)	This study
<i>T. roseus</i> , strain KBS 63	2,535 $\pm$ 0 (<i>SwaI</i> , n=1) 2,292 $\pm$ 352 (<i>PmeI</i> , n=2)	This study
<i>T. sp.</i> , strain TAA 43	2,069 $\pm$ 206 (<i>SwaI</i> , n=3) 2,457 $\pm$ 221 (<i>PmeI</i> , n=7)	This study
Subdivision 3		
" <i>S. usitatus</i> ", strain Ellin6076	9,967	Ward et al., 2009
Strain MPL3	5,869 $\pm$ 59 (<i>SwaI</i> , n=4) 5,829 $\pm$ 58 (<i>PmeI</i> , n=5)	This study
Subdivision 4		
" <i>Chloroacidobacterium sp.</i> "	ca. 3,500-4,000*	This study

*Premilinary pyrosequencing run done by the Joint Genome Institute, Los Alamos.

A. Ellin 6076



B. Ellin 345

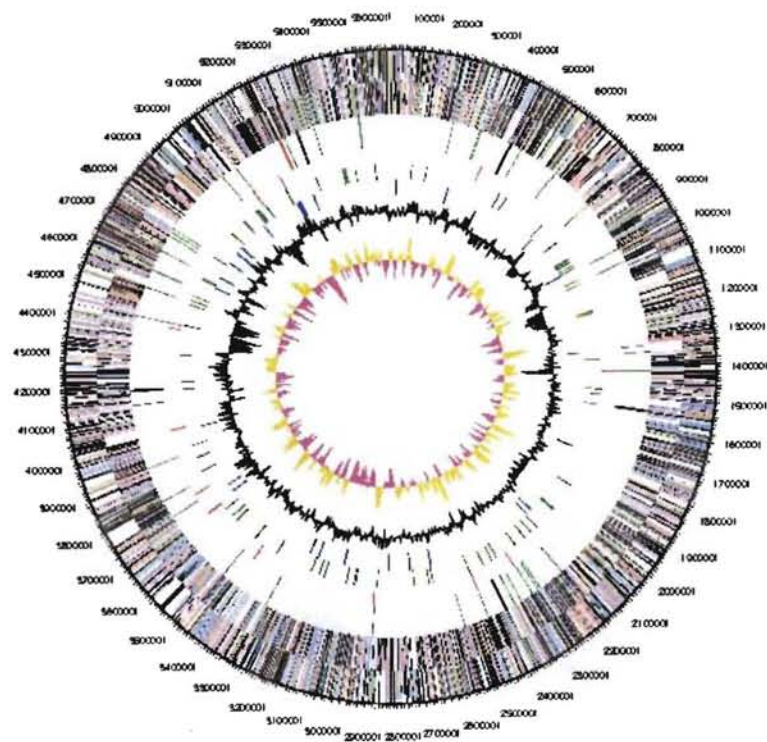


Figure 2

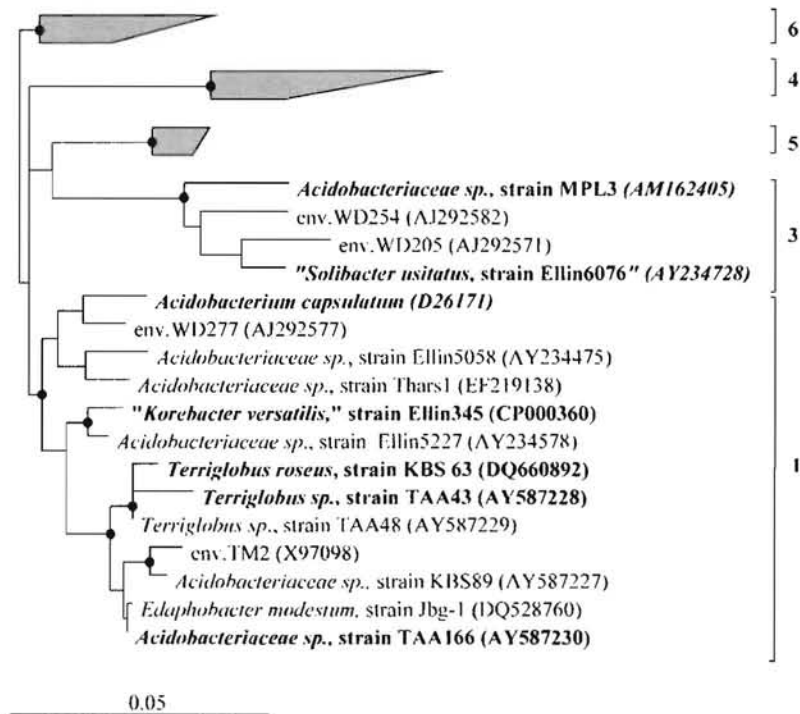


Figure S1

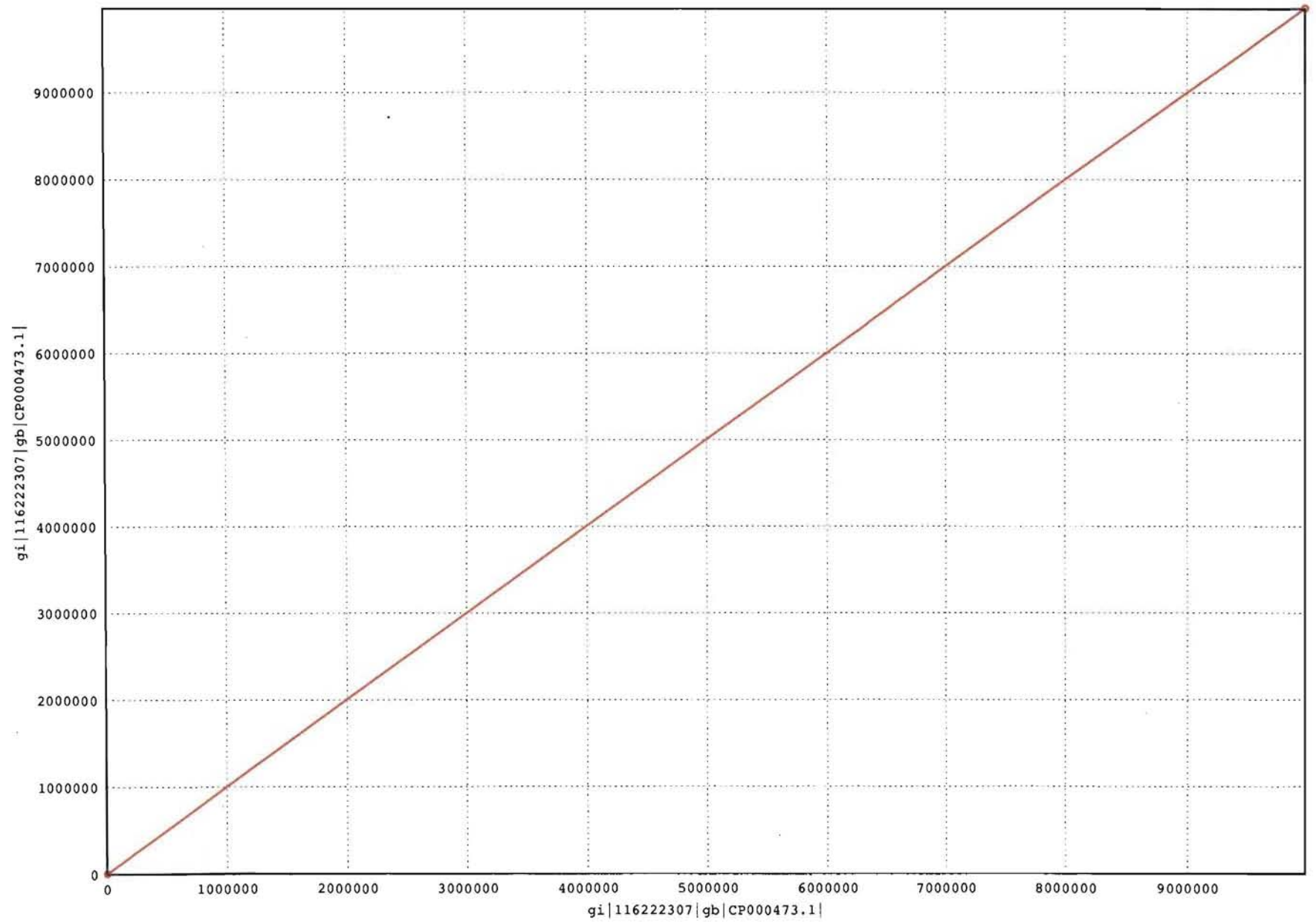


Table S1 is over 100 pages of output from a free program called repeat-match. The output looks just like that of Table S2. I didn't include it (Table S1) due to the length.

Table S2. exact-tandems output
 exact-tandems output for Ellin 6076
 Finding matches
 Tandem repeats

Start	Extent	UnitLen	Copies
76808	64	15	4.3
201645	48	24	2.0
297472	39	14	2.8
622440	31	8	3.9
1094140	92	36	2.6
1094305	158	72	2.2
1146712	42	16	2.6
1660794	77	36	2.1
1752151	107	45	2.4
1752217	143	45	3.2
2015276	37	14	2.6
2700665	56	19	2.9
2804765	54	23	2.3
2804866	54	24	2.2
2804905	74	23	3.2
2805148	114	48	2.4
2805311	52	22	2.4
3548924	26	3	8.7
3878932	41	15	2.7
4630516	53	21	2.5
4742596	40	18	2.2
4742780	90	33	2.7
5299146	37	12	3.1
5413935	325	150	2.2
5473806	36	14	2.6
5679332	45	17	2.6
5830824	38	16	2.4
6536004	53	18	2.9
7975920	34	14	2.4
8511223	60	28	2.1
8649960	548	261	2.1
9264063	32	8	4.0

exact-tandems output for Ellin 345
 Finding matches
 Tandem repeats

Start	Extent	UnitLen	Copies
1277502	76	19	4.0
1277539	76	37	2.1
1277558	38	18	2.1
1277576	43	19	2.3

1277601	75	19	3.9
1499595	58	29	2.0
1701738	50	17	2.9
1899855	94	40	2.4
1899870	187	80	2.3
2013936	32	7	4.6
2027213	47	19	2.5
3127358	41	16	2.6
3147967	36	12	3.0
3150267	47	23	2.0
3155379	365	100	3.6
3395523	47	22	2.1
3395578	61	23	2.7
3545753	42	13	3.2
3545783	35	13	2.7
4574596	86	42	2.0

Table S3. Tandem Repeats Finder output.

Tandem Repeats Finder Program written by:

Gary Benson

Department of Biomathematical Sciences

Mount Sinai School of Medicine

Version 4.00

Please cite: G. Benson, "Tandem repeats finder: a program to analyze DNA sequences" Nucleic Acid Research(1999) Vol. 27, No. 2, pp. 573-580.

Sequence: gi|116619145|ref|NC_008536.1| Solibacter usitatus Ellin6076, complete genome Parameters: 2 7 7 80 10 50 500 Length: 9965640

Tables: 1 2 3 This is table 1 of 3 (302 repeats found)

Click on indices to view alignment

[Table Explanation](#)

Indices	Period Size	Copy Number	Consensus Size	Percent Matches	Percent Indels
73203--73238	18	2.0	18	88	0
73244--73277	8	4.0	8	88	11
76793--76871	15	5.3	15	96	0
110100--110128	15	1.9	15	100	0
116751--116781	15	2.1	15	100	0
201626--201692	24	2.8	24	97	0
262319--262351	13	2.5	13	95	0
279490--279519	15	2.0	15	93	0
297472--297510	14	2.8	14	100	0
305190--305863	105	6.4	105	90	0
305190--305863	315	2.1	315	90	1
306510--306538	15	1.9	15	100	0
375576--375623	24	2.0	24	88	8
441732--441760	15	1.9	15	100	0
487163--487189	10	2.7	10	100	0
488842--488881	21	1.9	21	84	0
531494--531527	17	2.0	17	100	0
577772--577811	21	1.9	21	85	10
613123--613161	15	2.6	15	91	0
622440--622470	8	3.9	8	100	0
769688--769736	25	2.0	25	88	4
785043--786055	441	2.3	441	81	4
Indices	Period Size	Copy Number	Consensus Size	Percent Matches	Percent Indels
789675--790781	441	2.5	441	80	3
800936--802638	441	3.9	440	89	1
804812--804839	13	2.2	13	100	0
805353--805498	66	2.2	66	82	4
829913--831222	444	3.0	443	87	1
872579--872618	12	3.3	12	82	0
878216--878245	15	2.0	15	93	0
917255--917284	15	2.0	15	93	0
926219--926251	12	2.8	12	95	0
1094095--1094483	36	10.8	36	90	1

1123336--1123384	15	3.3	15	85	5
1129551--1129654	54	1.9	54	98	0
1146712--1146753	16	2.6	16	100	0
1180838--1180864	9	3.0	9	100	0
1230004--1230046	22	2.0	21	81	4
1253889--1253928	21	1.9	21	84	0
1253882--1253941	30	2.0	30	90	0
1312088--1312131	6	7.3	6	81	0
1312882--1312921	15	2.7	15	88	0
1316543--1316567	13	1.9	13	100	0
1371036--1371063	14	2.0	14	100	0
1511245--1511285	18	2.3	18	80	16
Indices	Period Size	Copy Number	Consensus Size	Percent Matches	Percent Indels
1553927--1553966	15	2.7	15	88	0
1638816--1638849	16	2.1	16	94	0
1654056--1654088	15	2.1	16	88	5
1660578--1660890	36	8.7	36	90	2
1660875--1661001	36	3.5	36	84	2
1660626--1660927	72	4.2	72	87	3
1660974--1661026	27	2.0	27	92	0
1661094--1661188	27	3.5	27	81	8
1702542--1702571	15	2.0	15	93	0
1723885--1723960	24	3.1	25	92	5
1752074--1752429	45	7.9	45	94	0
1811301--1811343	14	3.1	14	93	3
2015276--2015312	14	2.6	14	100	0
2087043--2087072	15	2.0	15	93	0
2114418--2114465	22	2.1	23	88	3
2153156--2153191	18	2.0	18	88	0
2199266--2199295	15	2.0	15	93	0
2431519--2431552	16	2.1	16	94	0
2457936--2457966	15	2.1	15	93	0
2529291--2529543	116	2.2	115	85	2
2593308--2596766	276	12.5	274	83	5
2623498--2623539	21	2.0	22	85	4
Indices	Period Size	Copy Number	Consensus Size	Percent Matches	Percent Indels
2700665--2700720	19	2.9	19	100	0
2804739--2805469	23	32.2	23	82	8
2805042--2805075	17	2.0	17	94	0
2804739--2805466	318	2.3	313	84	5
2884352--2884742	195	2.0	196	96	0
2962394--2962432	12	3.3	12	92	0
3031374--3031405	16	1.9	17	93	6
3037855--3037883	15	1.9	15	100	0
3272788--3272814	14	1.9	14	100	0
3274249--3274435	101	1.8	102	89	2
3352036--3352119	31	2.8	30	89	3

3352036--3352126	15	6.0	15	88	2
3352046--3352133	46	1.9	46	88	0
3388631--3388673	15	2.8	15	92	3
3424751--3424781	12	2.6	12	94	0
3477573--3477650	18	4.3	18	78	0
3477573--3477650	36	2.2	36	85	0
3548924--3548949	3	8.7	3	100	0
3622211--3622238	14	2.0	14	100	0
3659334--3659574	116	2.1	116	85	4
3739813--3739952	54	2.6	54	76	8
3768701--3768731	15	2.1	15	93	0
Indices	Period Size	Copy Number	Consensus Size	Percent Matches	Percent Indels
3878932--3878972	15	2.7	15	100	0
3888245--3888918	279	2.4	279	89	2
3904923--3904981	15	3.9	15	86	0
3949684--3949726	23	1.9	23	85	4
3960439--3960478	22	1.9	22	85	10
3966986--3967025	15	2.6	16	81	18
3969105--3969138	15	2.3	15	94	0
3976661--3976734	24	3.1	24	82	0
4042677--4042703	12	2.3	12	100	0
4101827--4101856	15	2.0	15	93	0
4111428--4111454	13	2.1	13	100	0
4271231--4271281	24	2.1	24	86	13
4271235--4271283	12	4.1	12	80	17
4271411--4271438	12	2.3	12	100	0
4357903--4357927	12	2.1	12	100	0
4366544--4366590	24	2.0	24	91	0
4366615--4366663	24	2.0	24	92	0
4542820--4542853	15	2.3	15	94	0
4582528--4582552	12	2.1	12	100	0
4584185--4584291	15	7.1	15	78	4
4592688--4592752	18	3.6	17	70	15
4592791--4592834	18	2.4	18	92	0
Indices	Period Size	Copy Number	Consensus Size	Percent Matches	Percent Indels
4592787--4592848	18	3.4	18	84	4
4605632--4605666	15	2.3	15	90	0
4605729--4605768	14	2.9	14	92	0
4608445--4608480	15	2.4	15	90	0
4630509--4630568	21	2.8	21	97	2
4681472--4681517	21	2.2	21	84	7
4688584--4688646	18	3.5	18	80	0
4697921--4697963	18	2.4	18	84	7
4724031--4724072	22	1.9	22	90	9
4742478--4742516	15	2.6	15	95	0
Indices	Period Size	Copy Number	Consensus Size	Percent Matches	Percent Indels

4742602--4742634	6	5.5	6	86	13
4742596--4742635	18	2.2	18	100	0
4742731--4742901	33	5.2	33	88	3
4742780--4742883	15	6.5	15	79	14
4742780--4742910	18	7.8	18	78	14
4742810--4742900	15	5.7	15	69	19
4813579--4813634	18	3.1	17	78	17
4815214--4815254	21	2.0	21	85	0
4815276--4815326	24	2.1	24	89	7
4815300--4815365	30	2.2	30	88	0
4851169--4851198	15	1.9	15	93	6
4876630--4876664	17	2.1	17	88	0
4945918--4945947	15	2.0	15	100	0
5025291--5025462	21	8.2	21	81	3
5025291--5025462	42	4.1	42	80	1
5025322--5025462	63	2.2	63	83	2
5026225--5027077	249	3.4	248	86	3
5031894--5031921	14	2.0	14	100	0
5034646--5034674	10	3.0	10	95	5
5142843--5142868	13	2.0	13	100	0
5208765--5208820	27	2.2	26	84	9
5223181--5223210	15	2.0	15	93	0
Indices	Period Size	Copy Number	Consensus Size	Percent Matches	Percent Indels
5278140--5278177	17	2.2	18	85	4
5299136--5299182	12	3.9	12	94	0
5299713--5299755	21	2.1	20	86	4
5299758--5299824	12	5.6	12	76	0
5299766--5299824	24	2.5	24	91	0
5324592--5324627	18	1.9	19	88	5
5376031--5376073	18	2.4	18	84	0
5402134--5402184	21	2.3	23	83	12
5413409--5413455	24	2.0	23	83	4
5413349--5413556	114	1.8	114	95	0
5413474--5413525	24	2.2	24	82	6
5413463--5413638	84	2.1	84	90	0
5413576--5413622	24	2.0	24	100	0
5413547--5413752	114	1.8	114	96	0
5413663--5413709	24	2.0	24	83	8
5413690--5413727	12	3.2	12	78	14
5413677--5413733	24	2.4	24	88	5
5413349--5413784	198	2.2	198	89	0
5413811--5414552	150	4.9	149	90	1
5414500--5414671	78	2.2	78	82	6
5467939--5468680	161	4.7	161	88	2
5473800--5473841	14	3.0	14	96	0
Indices	Period Size	Copy Number	Consensus Size	Percent Matches	Percent Indels
5554496--5554535	21	1.9	21	84	0

5659141--5659167	9	3.0	9	100	0
5679197--5679620	17	25.1	17	87	0
5723774--5723809	16	2.2	17	95	5
5726680--5726727	23	2.0	25	84	8
5736762--5736808	24	1.9	24	82	4
5759889--5759918	14	2.1	14	100	0
5803760--5803799	15	2.7	15	96	0
5803756--5803808	15	3.5	15	87	5
5830816--5830861	8	5.8	8	92	0
5830816--5830861	16	2.9	16	96	0
5830814--5830861	24	2.0	24	95	0
5858092--5858118	13	2.1	13	100	0
5890200--5890921	252	2.9	250	82	3
5890351--5891014	252	2.6	252	82	2
5925204--5925412	116	1.8	116	86	2
5997755--5997983	126	1.8	126	87	3
6026875--6026908	16	2.1	16	100	0
6052881--6052920	21	1.9	21	85	10
6059656--6059682	13	2.1	13	100	0
6107659--6107697	19	2.0	20	85	5
6154938--6154993	18	2.9	18	75	14
Indices	Period Size	Copy Number	Consensus Size	Percent Matches	Percent Indels
6169656--6169682	12	2.3	12	100	0
6185416--6185449	17	1.9	18	88	5
6207683--6207737	24	2.3	23	82	14
6402096--6402139	15	2.9	15	89	0
6438022--6438071	24	2.2	22	89	10
6446449--6446475	14	1.9	14	100	0
6467834--6467884	24	2.1	23	96	0
6467849--6467927	24	3.3	24	90	0
6516654--6516684	15	2.1	15	93	0
6535990--6536247	18	14.4	18	81	2
6676159--6676203	24	2.0	23	82	8
6770836--6770876	18	2.5	17	81	14
6771364--6771455	36	2.6	36	83	10
6771364--6771458	18	5.3	18	75	7
6771364--6771468	36	2.9	36	84	8
6771457--6771550	18	5.2	18	78	0
6771475--6771536	9	6.9	9	79	0
6771362--6771554	27	7.1	27	74	10
6771457--6771550	36	2.6	36	86	0
6785739--6785767	12	2.3	13	94	5
6802459--6802509	25	2.0	25	88	0
6802470--6802509	13	3.2	13	85	7
Indices	Period Size	Copy Number	Consensus Size	Percent Matches	Percent Indels
6802471--6802511	12	3.3	12	89	3
6855454--6855505	15	3.7	15	77	15

6855430--6855513	27	3.2	26	85	8
6855430--6855556	12	9.8	12	72	17
6855493--6855566	27	2.9	27	80	6
6855430--6855587	66	2.4	66	83	2
6930827--6931068	89	2.7	87	93	1
6957935--6957972	20	1.9	20	94	0
7057506--7057545	21	1.9	21	84	0
7179921--7179963	15	2.9	15	85	0
7185208--7185275	18	3.8	18	92	0
7228058--7228877	264	3.1	263	83	3
7236511--7236545	15	2.3	15	85	0
7250865--7250899	18	1.9	18	100	0
7375977--7376011	18	1.9	18	94	0
7376040--7376104	30	2.2	30	81	15
7376028--7376104	21	3.4	21	74	19
7376244--7376290	12	3.6	13	80	13
7395547--7395576	14	2.1	14	93	0
7398233--7398271	15	2.4	17	87	12
7489649--7489700	21	2.5	21	90	0
7525717--7525776	17	3.4	18	81	6
Indices	Period Size	Copy Number	Consensus Size	Percent Matches	Percent Indels
7582905--7584339	381	3.8	383	93	2
7589088--7589173	23	3.7	23	87	3
7590950--7591018	25	2.9	23	77	8
7666294--7666333	18	2.2	18	86	8
7708488--7708523	15	2.4	15	95	0
7718779--7718821	21	2.0	21	81	0
7809760--7809803	23	2.0	23	86	13
7826580--7826605	13	2.0	13	100	0
7829560--7829602	12	3.6	12	93	0
7930092--7930125	17	2.0	17	94	0
Indices	Period Size	Copy Number	Consensus Size	Percent Matches	Percent Indels
7975906--7975992	13	6.5	13	88	2
7980304--7980343	21	2.0	20	85	5
8009559--8009646	27	3.3	27	70	12
8009573--8009607	18	1.9	18	94	0
8018287--8018311	12	2.1	12	100	0
8054098--8054228	9	14.2	9	74	7
8054119--8054228	27	4.0	27	77	6
8054097--8054228	18	7.7	18	74	13
8054145--8054249	48	2.2	47	79	5
8055154--8055221	27	2.5	27	80	0
8055148--8055215	9	7.6	9	76	0
8055147--8055212	18	3.7	18	87	0
8055151--8055233	36	2.3	36	85	0
8055938--8055973	19	1.9	19	94	5
8061350--8061395	24	1.9	24	86	0

8099292--8100467	417	2.8	417	97	0
8198184--8198217	18	1.9	18	88	5
8208513--8208545	16	2.1	15	94	5
8250322--8250379	19	3.0	19	80	5
8296605--8296630	13	2.0	13	100	0
8353253--8353317	21	3.2	21	71	20
8353241--8353326	39	2.2	39	85	0

Indices	Period Size	Copy Number	Consensus Size	Percent Matches	Percent Indels
8409743--8409767	12	2.1	12	100	0
8446533--8446557	12	2.1	12	100	0
8461935--8461966	15	2.1	15	94	0
8490479--8490758	105	2.7	105	80	5
8490569--8490797	105	2.2	105	82	1
8510560--8510584	12	2.1	12	100	0
8511220--8511309	11	9.1	11	81	14
8511218--8511299	28	3.0	28	96	1
8534940--8535008	23	2.9	24	80	6
8541772--8543197	381	3.7	383	92	1
8649649--8653102	261	13.2	260	90	1
8760258--8760291	12	2.8	12	86	0
8768987--8769016	10	3.0	10	95	0
8780975--8781007	16	2.0	17	88	5
8793283--8793311	15	1.9	15	100	0
8798631--8798668	15	2.6	14	88	12
8818178--8818202	10	2.5	10	100	0
8961061--8961090	15	2.0	15	93	0
8972759--8972799	14	2.9	14	92	0
8982641--8982672	17	2.0	17	88	11
9034013--9034048	18	2.0	18	94	0
9038525--9038562	19	2.0	20	84	5

Indices	Period Size	Copy Number	Consensus Size	Percent Matches	Percent Indels
9264063--9264094	8	4.0	8	100	0
9447441--9447487	24	1.9	25	86	4
9554336--9554458	23	5.3	23	78	3
9554323--9554458	46	3.0	46	74	1
9557064--9557149	23	3.7	23	87	3
9589246--9589292	21	2.3	21	85	3
9602289--9602325	18	2.1	18	100	0
9636588--9636664	15	5.3	15	77	14
9636588--9636673	42	2.0	42	86	0
9636676--9636728	27	2.0	27	84	0
9691508--9691534	13	2.1	13	100	0
9770019--9770060	18	2.5	17	78	14
9770029--9770064	18	2.1	17	89	5
9770988--9771012	12	2.1	12	100	0
9870866--9870913	24	2.0	23	84	11
9886255--9886310	24	2.4	23	77	16

9941891--9941929	18	2.2	18	86	9
9955127--9955155	10	2.9	10	100	0

The End!

Tandem Repeats Finder Program written by:

Gary Benson
Department of Biomathematical Sciences
Mount Sinai School of Medicine
Version 4.00

Please cite: G. Benson, "Tandem repeats finder: a program to analyze DNA sequences" Nucleic Acid Research(1999) Vol. 27, No. 2, pp. 573-580.
Sequence: gi|94967031|ref|NC_008009.1| Acidobacteria bacterium
Ellin345, complete genome Parameters: 2 7 7 80 10 50 500 Length:
5650368

Tables: 1 [2](#) This is table 1 of 2 (154 repeats found)

Click on indices to view alignment

[Table Explanation](#)

Indices	Period Size	Copy Number	Consensus Size	Percent Matches	Percent Indels	Score	A	C	G	T	Entropy (0-2)
83876--83915	16	2.3	19	70	25	50	37	22	30	10	1.87
83881--83915	16	2.2	16	94	0	61	40	20	28	11	1.87
97390--97512	53	2.3	53	78	10	135	17	24	27	30	1.97
122824--122867	22	2.0	22	90	0	70	15	20	38	25	1.92
314316--314377	18	3.1	18	72	24	52	4	25	48	20	1.70
484065--484095	16	1.9	16	93	0	53	12	19	45	22	1.84
601921--602008	31	2.9	31	83	10	119	23	17	28	30	1.97
601921--602008	29	3.0	29	86	6	115	23	17	28	30	1.97
603246--603279	18	1.9	18	88	5	52	17	32	32	17	1.94
607986--608172	51	3.6	50	76	13	184	20	32	24	22	1.98
676899--676973	39	1.9	39	83	0	96	5	33	24	37	1.78
676971--677029	15	3.9	15	79	0	55	6	27	28	37	1.82
791899--791936	19	2.0	19	85	10	51	18	34	18	28	1.95
1067330--1067386	24	2.4	24	85	5	71	10	21	38	29	1.87
1067349--1067395	24	2.0	24	91	0	76	12	19	40	27	1.88
1101330--1101358	15	1.9	15	100	0	58	0	20	44	34	1.52
1277426--1277684	19	13.7	19	94	0	403	27	37	18	16	1.92

1296500--1296529	14	2.1	15	93	6	53	30	10	33	26	1.89
1499595--1499702	29	3.7	29	93	2	171	31	25	24	19	1.98
1500338--1500367	15	1.9	16	93	6	53	30	23	33	13	1.93
1515618--1515642	9	2.8	9	100	0	50	36	20	44	0	1.52
1556731--1556778	24	2.0	24	87	0	69	25	31	29	14	1.95
Indices	Period Size	Copy Number	Consensus Size	Percent Matches	Percent Indels	Score	A	C	G	T	Entropy (0-2)
1701625--1701787	17	9.6	17	88	2	206	29	22	27	20	1.98
1701626--1701789	34	4.8	34	90	4	251	29	21	27	21	1.98
1709050--1709080	16	1.9	16	100	0	62	16	32	32	19	1.94
1711164--1711243	27	3.1	25	80	10	81	5	23	53	17	1.63
1711173--1711261	27	3.7	24	76	12	76	6	24	50	17	1.70
1711233--1711294	21	3.0	21	80	4	72	12	22	48	16	1.80
1773814--1773851	18	2.1	19	85	5	51	28	34	26	10	1.90
1774041--1774085	21	2.1	21	83	0	54	11	33	44	11	1.75
1774283--1774323	21	2.0	21	85	0	55	12	31	41	14	1.83
1775201--1775243	18	2.4	18	81	14	61	44	6	46	2	1.43
1804645--1804763	53	2.2	54	85	4	170	26	21	25	26	2.00
1817553--1817593	12	3.4	12	79	0	55	0	29	36	34	1.58
1899850--1900062	40	5.3	40	97	0	390	15	39	24	19	1.91
1899969--1900258	153	1.9	154	91	0	483	16	37	24	22	1.94
1900122--1900216	40	2.4	40	89	0	136	18	35	25	20	1.95
1900936--1900978	21	2.0	21	95	0	77	6	58	25	9	1.54
1900936--1900990	18	2.8	20	83	8	69	10	56	21	10	1.64
1901090--1901137	21	2.3	21	96	0	87	18	41	16	22	1.90
1901101--1901153	21	2.5	21	87	0	70	20	37	16	24	1.93
1900985--1901274	157	1.8	158	96	0	537	15	40	22	21	1.91

1901111-- 1901660	137	4.0	136	91	3	867	17	38	21	22	1.93
2013936-- 2013972	7	5.3	7	96	0	65	59	29	0	10	1.31
Indices	Period Size	Copy Number	Consensus Size	Percent Matches	Percent Indels	Score	A	C	G	T	Entropy (0-2)
2027213-- 2027311	19	5.3	19	90	2	146	24	18	39	18	1.92
2027194-- 2027311	37	3.2	37	86	2	155	24	20	37	17	1.94
2074551-- 2074587	13	2.8	13	95	0	65	10	21	35	32	1.88
2081061-- 2081251	83	2.3	83	96	0	346	21	24	38	15	1.92
2275512-- 2275557	18	2.6	18	85	0	65	34	30	26	8	1.86
2383365-- 2383397	17	2.0	17	88	5	50	0	33	39	27	1.57
2394840-- 2394872	6	5.5	6	92	0	57	66	21	0	12	1.23
2394837-- 2394872	18	2.0	18	100	0	72	66	22	0	11	1.22
2445706-- 2445738	17	1.9	17	100	0	66	18	36	27	18	1.94
2496819-- 2496865	24	2.0	24	86	0	67	25	8	55	10	1.62
2519051-- 2519389	36	9.5	35	80	8	369	14	31	32	21	1.93
2519074-- 2519389	72	4.4	71	92	2	474	13	32	32	21	1.92
2529004-- 2529046	15	3.0	14	86	13	61	44	16	39	0	1.48
2766552-- 2766613	23	2.7	23	82	0	79	8	27	22	41	1.82
2795933-- 2795972	18	2.2	19	86	4	55	20	30	40	10	1.85
2829027-- 2829073	21	2.2	21	88	7	69	6	27	38	27	1.81
2829804-- 2829883	40	2.0	41	92	2	135	12	37	25	25	1.91
2829803-- 2829905	41	2.5	40	85	1	134	12	36	28	22	1.91
2847857-- 2847902	15	3.1	15	93	6	76	19	26	26	28	1.99
2958576-- 2958607	15	2.1	16	94	5	57	37	28	31	3	1.73
2999085-- 2999113	15	1.9	15	100	0	58	13	48	31	6	1.69
3002861-- 3002974	39	3.0	38	85	5	158	22	21	30	25	1.99
Indices	Period Size	Copy Number	Consensus Size	Percent Matches	Percent Indels	Score	A	C	G	T	Entropy (0-2)

3002875-- 3002981	39	2.8	38	87	2	160	24	22	30	22	1.99
3014931-- 3014970	18	2.2	19	86	4	55	20	27	12	40	1.88
3021632-- 3021997	93	3.9	94	79	11	439	15	40	23	20	1.90
3127358-- 3127398	16	2.6	16	100	0	82	4	39	36	19	1.73
3141023-- 3141090	19	3.6	19	87	0	100	17	26	26	29	1.98
3147956-- 3148002	12	3.9	12	94	0	76	21	17	36	25	1.94
3148059-- 3148175	54	2.1	55	90	4	184	29	34	23	12	1.92
3150262-- 3150313	23	2.2	23	96	3	95	19	53	9	17	1.70
3155211-- 3155265	20	2.8	20	81	8	58	9	29	21	40	1.84
3155253-- 3155740	51	9.7	51	87	3	704	19	31	26	22	1.98
3155253-- 3155743	100	4.9	100	96	1	874	19	32	25	22	1.97
3290118-- 3290176	12	5.3	12	75	20	65	35	13	42	8	1.75
3290122-- 3290175	23	2.5	22	90	6	83	38	12	38	9	1.76
3290122-- 3290177	10	5.2	10	80	20	60	39	12	39	8	1.75
3298304-- 3298362	12	4.8	12	80	12	66	20	37	20	22	1.95
3298341-- 3298368	12	2.3	12	100	0	56	21	32	25	21	1.98
3298348-- 3298387	12	3.3	12	89	0	53	22	22	27	27	1.99
3310179-- 3310212	17	1.9	17	88	5	50	20	23	44	11	1.84
3341726-- 3341824	43	2.3	44	87	5	141	24	26	31	18	1.97
3395523-- 3395660	23	6.1	23	96	0	237	41	21	23	14	1.89
3430093-- 3430266	55	3.1	55	91	4	262	16	26	25	32	1.96
3430136-- 3430226	18	5.1	17	72	17	78	15	27	25	31	1.95
Indices	Period Size	Copy Number	Consensus Size	Percent Matches	Percent Indels	Score	A	C	G	T	Entropy (0-2)
3432615-- 3432788	81	2.1	82	95	1	314	14	44	20	19	1.86
3434559-- 3434740	18	10.2	17	76	10	159	11	45	18	24	1.82
3434595-- 3434740	35	4.1	35	88	5	188	11	45	17	24	1.82

3525930-- 3526094	83	2.0	83	85	2	224	17	26	38	18	1.92
3532108-- 3532346	92	2.6	93	91	4	385	19	25	38	16	1.93
3539657-- 3539781	53	2.4	53	82	5	155	24	27	22	26	2.00
3545753-- 3545817	13	5.0	13	98	0	112	30	18	35	15	1.92
3671253-- 3671485	86	2.7	86	89	2	324	22	22	32	22	1.98
3773256-- 3773340	33	2.6	32	89	9	127	7	37	18	36	1.78
3851358-- 3851482	53	2.4	52	89	4	198	20	28	25	26	1.99
3855673-- 3855786	40	2.9	38	89	8	135	24	28	24	21	1.99
3855693-- 3855786	40	2.3	41	98	1	181	25	29	23	21	1.99
3978197-- 3978231	18	1.9	18	88	0	52	20	17	42	20	1.89
4109380-- 4109410	16	2.0	16	93	6	55	22	12	38	25	1.90
4139697-- 4139734	18	2.1	19	85	5	51	18	15	39	26	1.91
4149899-- 4149946	22	2.3	20	78	7	51	22	52	14	10	1.72
4228068-- 4228108	11	3.7	11	86	0	55	31	17	48	2	1.60
4228078-- 4228124	21	2.3	20	82	10	58	31	14	46	6	1.70
4228090-- 4228114	12	2.1	12	100	0	50	32	8	52	8	1.60
4233553-- 4234132	279	2.1	279	85	2	779	17	25	33	23	1.97
4251746-- 4251781	18	1.9	19	88	5	56	30	30	27	11	1.91
4301541-- 4301570	14	2.1	14	93	0	51	16	6	43	33	1.74
Indices	Period Size	Copy Number	Consensus Size	Percent Matches	Percent Indels	Score	A	C	G	T	Entropy (0-2)
4307335-- 4307376	18	2.3	18	87	0	57	35	30	16	16	1.92
4395568-- 4395593	12	2.2	12	100	0	52	50	34	7	7	1.60
4445912-- 4445936	11	2.3	11	100	0	50	0	40	28	32	1.57
4574368-- 4574459	42	2.2	42	84	0	112	20	26	35	17	1.95
4574542-- 4574801	42	6.2	42	83	0	241	18	29	38	14	1.90
4574542-- 4574801	84	3.1	84	82	0	268	18	29	38	14	1.90

4574548-- 4574783	126	1.9	126	86	0	337	18	29	37	15	1.91
4591804-- 4591837	15	2.3	15	94	0	59	0	14	35	50	1.44
4599571-- 4599644	21	3.5	20	83	3	76	6	39	18	35	1.78
4599571-- 4599653	42	2.0	42	83	4	105	7	39	19	33	1.79
Indices	Period Size	Copy Number	Consensus Size	Percent Matches	Percent Indels	Score	A	C	G	T	Entropy (0-2)
4600188-- 4600220	12	2.8	12	90	0	57	15	6	33	45	1.70
4637641-- 4637670	12	2.5	12	100	0	60	56	20	23	0	1.42
4683698-- 4683740	18	2.4	17	85	11	59	27	27	27	16	1.97
4683707-- 4683741	18	1.9	18	94	0	61	31	25	25	17	1.97
4693937-- 4694068	48	2.8	48	89	2	203	14	26	30	28	1.95
4709766-- 4709800	18	1.9	18	88	0	52	11	40	28	20	1.87
4735656-- 4735690	18	1.9	18	88	0	52	5	31	31	31	1.81
4743935-- 4743963	15	1.9	15	100	0	58	27	41	20	10	1.85
4744276-- 4744328	21	2.3	24	81	9	67	11	22	50	15	1.75
4826467-- 4826497	15	2.0	16	93	6	55	6	19	58	16	1.59
4827957-- 4828010	24	2.3	24	87	6	74	29	37	27	5	1.80
4828478-- 4828526	18	2.7	18	77	0	62	0	26	55	18	1.43
4922124-- 4922400	136	2.0	135	91	3	450	25	21	32	19	1.97
4929080-- 4929110	15	2.1	15	100	0	62	45	12	6	35	1.68
4943504-- 4943533	15	2.0	15	100	0	60	26	33	20	20	1.97
5065866-- 5065897	16	2.0	16	100	0	64	18	43	18	18	1.88
5122698-- 5122739	18	2.4	18	84	7	52	30	30	19	19	1.96
5156690-- 5156872	26	7.1	26	86	4	228	16	32	22	27	1.96
5156690-- 5156872	51	3.6	52	83	4	237	16	32	22	27	1.96
5156690-- 5156872	77	2.4	77	88	1	269	16	32	22	27	1.96
5173085-- 5173118	15	2.3	15	94	0	59	23	52	17	5	1.66

5189420--5189549	52	2.5	53	95	5	239	26	28	23	21	1.99
Indices	Period Size	Copy Number	Consensus Size	Percent Matches	Percent Indels	Score	A	C	G	T	Entropy (0-2)
5225928--5225958	15	2.1	15	93	0	53	38	9	35	16	1.81
5242882--5242916	18	1.9	18	88	0	52	8	34	34	22	1.85
5296354--5296408	24	2.4	23	85	8	69	10	7	32	49	1.65
5398849--5398927	25	3.2	25	75	13	99	11	34	29	25	1.91
5398825--5398976	50	3.0	50	88	0	214	15	30	25	28	1.96
5398859--5398943	25	3.4	25	78	12	111	12	34	28	24	1.92
5415529--5415621	24	3.9	24	92	0	159	29	35	19	16	1.93
5476028--5476203	59	3.0	59	95	0	316	20	32	28	18	1.96
5525289--5525321	17	1.9	17	93	0	57	15	27	9	48	1.74
5529559--5529703	32	4.6	31	87	4	193	30	23	26	19	1.98
5529559--5529703	63	2.3	62	84	4	193	30	23	26	19	1.98
5540511--5540620	34	3.2	35	89	5	172	26	33	18	21	1.96

The End!

Table S4. Inverted Repeats Finder output

Inverted Repeats Finder Program written by:

Gary Benson
Department of Biomathematical Sciences
Mount Sinai School of Medicine
Version 3.05

Sequence: gil116619145|ref|NC_008536.1| *Solibacter usitatus* Ellin6076,
complete genome

Parameters: 2 3 5 80 10 40 500000 10000 -d -h -t4 74 -t5 493 -t7 10000

```
35228 35261 34 35284 35317 34 22 94.1176 0.0000 58 44.1176 55.8824 43.7500
56.2500 0.0000 70545 70545 TCTCGCAGAGGACGCGGAGATCGCAGAGAAAACC
GGTTTTCTTTGCGATCTCTGCGTCTCTGCGAGA
73061 73107 47 73123 73169 47 15 80.8511 0.0000 49 54.2553 45.7447 55.2632
44.7368 0.0000 146230 146230
AAACACATTTCTCGCAGAGACGCGAGACGCGAGAAAACCGCAAAGAA
TTCTTTGCGTTACTTTGCGGCTTTGCGCCTTTGCGAGAAATGTCCTT
73060 73161 102 73192 73290 99 30 82.3529 2.9412 108 51.2438 48.7562
51.1905 48.8095 0.0000 146353 146350
CAAACACATTTCTCGCAGAGACGCGAGACGCGAGAAAACCGCAAAGAAAACCTGAAGTCTTTCTTCTTTGCGTTA
CTTTGCGGCTTTGCGCCTTTGCGAGAA
TTCTCGCAAAGCCGCAAAGACGCGAGAGAACCGCAAAGAAGAAGAGAAAGTTTTCTCTGCGTTACTCTGCGTCTCT
GCGTCTCTGCGAGAACCGTCTTTG
73185 73230 46 73243 73288 46 12 82.6087 0.0000 52 47.8261 52.1739 47.3684
52.6316 0.0000 146473 146473 AAAACCGTTCTCGCAAAGCCGCAAAGACGCGAGAGAACCGCAAAGAA
TTCTCTGCGTTACTCTGCGTCTCTGCGTCTCTGCGAGAACCGTCTT
100543 100581 39 100611 100649 39 29 82.0513 0.0000 43 35.8974 64.1026
37.5000 62.5000 0.0000 201192 201192
CGCCATCGGAGCCGCGACCGTGAGGGAGCGGTTGCTAAT
ATTACCGACCGCTCCTTAACAGTCGCGGCTCGCATGGCG
161405 161448 44 161476 161519 44 27 81.8182 0.0000 48 54.5455 45.4545
58.3333 41.6667 0.0000 322924 322924
TTTTCTCGCAAAGCTGCAAAGCCGCAAAGTTTCGCCAAGAACAA
TTTTTGTTGCGTTACTTTGCGCCTTTGCGCCTTTGCGAGAAAA
165844 165885 42 165898 165938 41 12 80.9524 2.3810 42 46.9880 53.0120
44.1176 55.8824 0.0000 331783 331782
TTCTCGCAGAGACGCGAGCCGCGAGAGAAGCGCCAAGAAGAA
TTCTCTTTGCGTTACTTTGCGGCTTTGCGGCTTTGCGAGAA
165901 165938 38 166000 166037 38 61 92.1053 0.0000 61 47.3684 52.6316
45.7143 54.2857 0.0000 331938 331938 TCTTTGCGTTACTTTGCGGCTTTGCGGCTTTGCGAGAA
TTCTCGCAAAGCCGCAAAGCCGCAAAGAACCGCCAAGA
```


165844 165885 42 166068 166109 42 182 85.7143 0.0000 54 42.8571 57.1429
44.4444 55.5556 0.0000 331953 331953
TTCTCGCAGAGACGCAGAGCCGAGAGAAGCGCCAAGAAGAA
TTTTTCTCCGCGGTTCTCTGCGTCTCTGCGGCTCTGCGAGAA
165997 166044 48 166065 166112 48 20 81.2500 0.0000 51 46.8750 53.1250
48.7179 51.2821 0.0000 332109 332109
CTGTTCTCGCAAAGCCGCAAAGCCGCAAAGAACCGCCAAGAACAAGAA
TTGTTTTTCTCCGCGGTTCTCTGCGTCTCTGCGGCTCTGCGAGAAAAG
394698 394726 29 394745 394773 29 18 96.5517 0.0000 53 43.1034 56.8966
42.8571 57.1429 0.0000 789471 789471 CGCAGAGCCGAGAGTCACGCAGAGAAAA
TTTTCTCTGCGTGACTCTGCGGCTTTGCG
462014 462083 70 462095 462163 69 11 78.5714 1.4286 63 32.3741 67.6259
32.7273 67.2727 0.0000 924178 924178
GTAGCGTTTGGTCCGCTCGATCCGATCGGCTCCATCCGAGCCGCGACCGTCAGGGAGCGGTCGCTAAT
ATTACCGACCACTCCCTGACGGTCGCGACTCCGATGGCGGCACATCGACGGGGCGCAGCCGGCGCTAC
487158 487200 43 487206 487248 43 5 93.0233 0.0000 71 54.6512 45.3488
55.0000 45.0000 0.0000 974406 974406
TGGGCCGAGATGAACGCAGATGAACGCAGATAAGAGAAAAAA
TTTTTGCTTTTATCTGCGTTTCATCTGCGTTTATCTGCGGCCCA
510788 510820 33 510865 510897 33 44 90.9091 0.0000 51 43.9394 56.0606
43.3333 56.6667 0.0000 1021685 1021685 TTCTCGCAGAGCCGAGAGTCACGCAGAGAAGA
TCTTCTTGCGGTTACTCTGCGGCTCTGCGAGAA
518975 519013 39 519029 519067 39 15 87.1795 0.0000 53 46.1538 53.8462
44.1176 55.8824 0.0000 1038042 1038042
TTTCTCGCAGAGGCCGAGAGATCGCAGAGAAAACCAAG
CTCGGTTCTCTCTGCGTTCTCTGCGGTTTCTGCGAGAAA
527223 527260 38 527317 527354 38 56 81.5789 0.0000 41 44.7368 55.2632
45.1613 54.8387 0.0000 1054577 1054577
CTTGCGCCACTTTGCGGCTTTGCGGCTTTGCGAGAAA
TTTGTCGAGAGACGCAAAGACGCAGAGTAACGCCAAG
590828 590862 35 590881 590915 35 18 85.7143 0.0000 45 31.4286 68.5714
26.6667 73.3333 0.0000 1181743 1181743 CACTCCCAACGAGCCGCGACCGTTAGGGGGCGCTA
TAGCGCTCCCTAGCGGTCGCGGCTCGTAAGGCGTG
622250 622295 46 622313 622358 46 17 86.9565 0.0000 62 48.9130 51.0870
47.5000 52.5000 0.0000 1244608 1244608
AAAACGGTTCTCGCAAAGCCGAGAGCCGCAAAGTAACGCAAAGAA
TTCTTGCGGTTCTTTGCGGCTTTGCGGCTTTGCGAGAACAGCTTT
622256 622351 96 622436 622541 106 84 77.7778 12.9630 68 46.5347 53.4653
47.6190 52.3810 0.0000 1244787 1244797
GTTCTCGCAAAGCCGAGAGCCGCAAAGTAACGCAAAGAAAGCCTAGAACGGATGTCTTCTTGCGGTTCTTTGC
GGCTTTGCGGCTTTGCGAGAA
TTCTCGCAAAGCCGCAAAGCCGCAAAGTGACGCAAAGAGAGCCCAAACAATTCTGCTTTCTTTGCG
TACTCTGCGGCTCTGTGCCTCTGCGAGAAC
622448 622485 38 622499 622536 38 13 81.5789 0.0000 41 42.1053 57.8947
41.9355 58.0645 0.0000 1244984 1244984
CGCAAAGCCGCAAAGCCGCAAAGTGACGCAAAGAGAGC

GCTTTCTTTGCGTTACTCTGCGGCTCTGTGCCTCTGCG
633441 633482 42 633503 633544 42 20 88.0952 0.0000 59 39.2857 60.7143
37.8378 62.1622 0.0000 1266985 1266985
CCTTTCTCGCAGAGACGCAGAGCCGCCGAGTAACGCCAAGAA
TTCTCTGCGTGACTCTGCGGCTCTGCGTCTCTGCGAGAACGG
675427 679862 4436 679997 684340 4344 134 67.1315 5.6662 1099 38.2346
61.7654 37.2239 62.7761 0.0000 1359859 1359767
GTCTACCAGCAGGCGTTTGGCCAGTGCGCTAAAACCTGAATTCGGTCACGACGCGATGCTGCAGATAACTGGCGAT
GTCGCGGCGCCGATGAGGCTTGGCCAGGTAGGCGTCCACCAGAGAATTAGATGGTCGATACCGCGGTACATCAC
TTGCGCGGCCGCATCGCCCATGATCCGGGCGAAATCTTCTTGTAAATCGAAGAGCACCAGGCCGCCGGCGGCGAA
GCAACCCATGATCTTGGGAGAAACGCCGGTGTGTAGCCAGGTTGACGACGTCCACAATTACGGAGCTGTTGAT
GAACAGCAGCGGCAATTCGGTGAAATAGTCGAGGGACTTGAGAAAACGGATGCCGTACTGCTTGCTTAGGCTGGC
AACGCTTCCGGGTTCCATGAAGTTGCCGTAGAAATCGTACTCGCGCTTCAGGCCCGTCAGGACGGCGAGGCGCAC
CTCGGTATTGGCCACGAGTTCGATCTCTTCGTGCATGAAGTTCCAGAAGAAAGTGGAGTCCGGCTCCAGACTCAA
CTGCTTGCGCGTCTTGCTATCGAGGGCCTGGATCTCCGCCACCAGGAGTCCAGGCACACTCGGTGAGGCGAGC
CTTCTTGCTTCCATGACTCGCGATTCCAGCGAGCCGAGGAGCGGTTGGTTGCGGAATGGGAGGGCGTGCGCGGC
CTTCAGATACAGTTGCCCGCAAAAGCAACCGTACTGCGAAACATGTTGCGGGGCACGGTGCGATATCCGTATTT
GACAAAGTTGGGGTAGGCCGGCTGCAAGAAGAATCGCACCTTGCCGGCTTCAGAATGCCGGCCTTATCGAGCGC
CGCGATATGACCTTGTGCGGAGAATAATGCAGGTACAGGGGATGATTGAGAACCTGGCGGATGCGGCTCACGGC
GCCTGTCCGCGCATCTTCCGGCGTGCTGGGTTGTGATTCAAATCTGGCGTGGAAGCTGGAGCAATCCGTGATC
CCACAGCATCAGCGTGGGTACGCACAGAATGTCGCGGAAGATATTCTGCTGGTCGCGCGTGACGCAAAGCAGCAC
GTAGAGGGCGTTGGGAAGCGAAATGACCACGTCCGGCCGAAAGCTCTTGAGGGCTTCGATGACGCCGGCGATGTA
ACTCTCGTTCTGTCCCGTGAAGCCGGCCACGTCCACGTGGAAGACCTTGGCGCCGAGCGCGCGCAGTTCTTCTGA
CATCTCGATGGCGAAGTTGTAGTCCACGCCTCCGAACCAGGCGTGCGGAGGCAAGGTTACGGCGATGCGGAGTCC
GGCCAGCGGAAGGCCTTCATCGCGAAGCACGCTCATGAAGAGTACCTCTTCGCCAAAGTGCGGACGAGATGCGCG
GCGTACCGAGGCGAGGCCGAGTATGCGCTGGTACATCGGACGCTCGCCGCCGTGGCCTTGAATTTGGGGATCGCA
ACCTCACGGCTTCCAATCACTACGCGCGCCAATAAGAAGCCTAGCGGCCGCGGATCGGCGTTCCAGCCAAGCGAT
TCCGGAGAGCGCGGTTGAGGGATTTCAAACACGAGAGTGAGCTCGGGCGCTGCCGCGGCGCCTTGAGCGAGGACG
GCCGCCGGTATGGTGGCGAACTGGCGGTGCGGATTGCGGTGCGCGAGCAGCCATTCGGCGATTGCCTGTCCGTTG
GCCTTGACGGTCACGGGCAACGGCTTGGCGCGCGCAGGGATCATGCAATCGGAGATCACGAACGCGCACGGCAGA
TCGCCCCGCGACCGGTCCATCGAAGGTACGTGGAAGCTGCAGCGCGTGCCGATGGTCCACGATCCGCGCGGGCCG
GGAAAGGACCACTGCGAGCCCATCGCATCGCCAACGAATGCCATGGAATCTCCGCCTTCGAAGAAATCGATGGGG
TCTCCCGGACGGTAGGTGAGTCGTCCGACCGGAGAGATGCGCAAGCGGCAGAGGCGGAATCCGAGAGGTGCGGTG
TCCCACGTTGACAGGCCAGGTGAGCGGCCTGCGCGGCGCTGCGATTTGGAAGGCGATGTTACGGGTTTGGAG
GGAAGTAGGGCATCGGGCGGAAGGAGGATGCGAAGTTCTCCGCGTCTCGAGCCGGCCCCAGGTTCCATTGGGCC
GCGGGAAGACGTTCCAGAGAACGCGCACGGTCTGCGTGGGGTGCTCGTTGTTGAGGGCGGCGTCTGTTGACGGTA
AAGACAACCTGCGCGCGCTGGTCCATGGGCTTATCCGGCAGCAGCGTCAATGAACTTCGAGGACCAAAGCTCCAC
ATTCCTCAAACCGTCGGGGGCCGACCAGCATTATCGGTGTGTACGGCGCCGGTGCCCGCAGCCGTGAAATCCAGT
ATCTGCCCGGGGCGGTAACGCAGGCGTTGGTAGCTGCGGAAGACAATCCTGTGGACTTTGAATCCGAGTGCGCGT
GGGTGTCGCGCAATACCTGCTCTCCCCTGGCTTTGGCCAGTCGCGCGGCGGATTGAGGATTCGGACATGGAAC
TCGAGACAGCAAATCTTTTCCAGCGAGGGAGGCCGCGGGGATGGTAACCTTGGCGTCCACTACTTCGTAGCGC
GAAAAGAAGCTCCAATTGGCGAGAGGGACTCCGTCCAGAGAAACCTCGACCCGATTGGGTGCGCTTCCGGTCCG
ATCAGCGCTTCGTTACGGTCACGCAGGCGACGACATCGCTTCGACCGGGGATGGCGGGCGAAAGAGTAGGACA
GCGCGGTGCGCATCGGTCCAGGTGCCGGTGTGATCGGTTGCTCCCAATGGCCGTCTTTATGCGATGCGGCCGGG
CCGGCGGAGGTGAAATCGATGCTCTCATCGATGGCGACGGGCGGCACGGAAGCATCGTCAACGCCGAAGCCAAAG
AAAGAGCGCATATCGTCTTCGGTCTGCTTCCAGGTATTGGAGCGATAGCGGCGCTCCTCGCCACGCGCGGCCTT

CCTTTGCGTACGTGGGTAGCGAACGGCGTGAAAGTAAGGAGCAGGAACTGGTTACGAAATTGCGGAACGTGTGG
TGATTGGCGTTGATGACGACGACGTCCGGGACGCGCGTGCCGCCCCGACCAGCCAACTCGGCGCACATTTTCGAGA
CAGCGAGCCACGGCGCCAGGCTGTAGCGCAGAGGGCGCGGCCAGTCGGTTACACGCTCCGCCGGCATTCCGTTG
AGCCCCCTCCACGTCGAGGCTGAGATCGAAGACTAACAAGTCGTAGCTTGCGAGCAGCGTTTCGGGGAGCGGGTCG
GCATCGATCGCGCGGGCGTTTTTCGGGCGGAGCTCCGCTGGTCAGCAGCGATGCGGCATAGTCCAAAGGGGTTTCA
AACCCCAACTCCACGACACAGCGGGCGGCGAGCGACAGCATGCGTGATGATTGCCGACATACGCAAAACGCGCG
GAGCGGGTGAATACCGTCAGATCGTTGGCTAGGTATGGCAGTACATGCTCCGCCTGCACCGATAACGAAATTTTCGG
ATGCCGACGGCGTCGGACTGGTAGTACGACTCCTGCGGCGCGCCCAACGTGCGCTCCACCGCGGTGACGAATCGG
GCCTGCGGGTCGTTGGCAAGATCCAGCTCTTCAATCGGTTTCGTTGGGCAGACCCCAGGTTTCGGCCTGGTGGCGT
GCGTAGGGGTCCTCCAGGGAGAACACGAACTCCTGCACGTTGTTGTCGAGCTTCATGTCGAACTGGTGCATTCCG
GTCTGCACGCGGGTATGGTCGAGTGGTATCCCTTTAAGCGGTGCGCCAAGCTCTCGGTTGACCGCCGTAATAG
ATGTAGAGACGCTTGACGAGCTTGAGTACAGTGCCAGCCGTGAATCATGCGTTCGTCGAAGCCGTGAATATCC
CATAGGGCCTGGCGCGGCGACAGCATGAAATCGCCCACTTGATCGAAGCGGCTTTTCGGGAAAAGAGAGGGTCACT
TCATCCAGGTGCAGCCGCGGACCCAGATCGCGGCGAGGCTTGAGAGTGGCTTGCGGATCGCTGCGCGGGAAGGTC
TCCAGAGCGCCTCCGGTAGTTCAAACCTAGGGAGGGTGATAAGCCGTGCGACAGGTGCGCGCGACATCCGTC
AGGTCCCGGAAGCCGGGGCGCGGCACAAGGAGGATGTCCGTGTTGGTGTTCAGGACCCAATGATTGCGCGGGTTG
GAGCGACGACGCGGATATTCCGCGAGGCGGTTGACCAACCATCAGGTGAGTGTAGCCACCGACCAACGCTTA
TGAAGCTCCGGACGAACGCGAAAGACGCGCAGGAGGGCTTTGGCCCGGGGTGTCAGGGTATCGTAGATGGCTTCG
ACGAAAGTGGGGAGTTTCGTTGGACGTATTGCAATCGACGAAGAGAATTTCTGTCGTCGGATCCGACAACAATGCG
GCGAGACAATTCAGGCTGATTGCGCCGCGCTTGTCAGATTATAACCGTAAGAATCGTTGCGTCCGTAGAGGATT
ACCGAGATCAT

ATGATCTCGGTTATTCTCTACGGGCGAAACGATTTCGATGGTTACAACCTGCACAAGCGCGCGGCCATTAGCCTG
AATTGCATCGCGGAAGTGCTCAGCGATCCTGATGACGAGATCCTCTTCGTTGACTACAATACGCCCAACGACCTG
CCCACCTTTGTCGAGGCCATCTACGACACACTGACGGAGGCCGTCAAATCACGGCTGCGCGTCTTCCGCGTCCGT
CCGGAAGTTCATCAGCGGCTCGCCGGAAGCACGCATCTCGCGGCGCTAGAACCATTTCGCGGAACATCGCGATT
CGCCGCTCCAATCCGCGCAACCGCTGGGTCTGCTACCAACACGGACATGATCTTTCTGCCCCGCGCGAGTTAC
ACCACCTCTCCGGCGCCGTAGGAGATCTTGCCGACGGTCTCTATATCGTTCCGCGCTATGAGCTTCCCGAGCCG
CTCTGGGAGGCCTTCCCGCGCACCGACCCGAGTCGGTCTGCGGGCCTGCGAAGATCTCGGCCGGAACCTCCAC
CTGGAAGAGATTACGCTGTCTCTTCTGAAATGCGGTTGATCAGCCTGGCGATTTCCAGTTGGTGCCCCGCCAG
GCGCTTTGGGATATCAATGGCTTCGATGAGCGAATGATCCACGGCTGGCACGCCGATGCCAACATATGCAAGCGT
TTCTTTTTGTTTTATGGCAACCGCACGGAAGCCTCGCCGATCGGGTGAAAGGCTTTCACTGCGATCACACCCGC
GTTGCGACGCTCGCCATCGCCTGGACATCAAGCTCGAAAAACAACCTCAACCAATTCTTCTGACCTGAAAGAC
CCGGTCGCGCACCAACAGGCCGAAACCTGGGGCGCGCCCGGCGAAGAGATTGAAGAGCTGAACTTCCAGGACGAC
CCGCCGGCGCGCTACGTTTCGGCGCTGCGCGGCGTCATCGGCCCCGCCAGCCAGCCGAATATGTGGCCGATTCC
AACGACCTTCGAATTTCTGTCGCGTACCACGCCGAGCACGTCTGCCGCACCTCTGCGGATGTCTCACGACGTAT
CCCCGCGCTGCCCCGATTGCTACATCGGCGCAATCCGCACATGCGCGGGTTGAGCACGCGCGCCATCGCCGGA
CTTGCTTCAGCGAGCCCTTGTTGACGCCGGCGTCGCCGCCGACCCGAGACCCTGCTCGCCGGCTATGACCTG
CTGCTCTTCGATTTGCGCCTGGACCCAGCCGACAACCCCTGGATACCGTAGCGCGGGTCACCGATTGGCCGCGA
GAGCTGCGCTACCGGCTCGGCGACGCCGCCAAATTGCTGGAGGCGTGCGCGCTGCTCGCCGGCCAGAATGGCTCC
CGAGCCCGGACTTTCTTACCATCAACGCCAATCACACATCTTCCGCCAGTTCGCCGGCCAGTTCCTGTTGATG
CCCGAAACGCCCTACGCCACTCACGTTTCGAAAGGCCCGCCGCGGTGGGCGAGGAGCGTCTTACCGGGGGCCAC
ACCTGGAAATACGTGGAAGACGACATGCGCTCGTTCTTCGGCTACGGCAAATGGGAGAACACCATTTCCGCGCATC
ACGGCCGGCGACTCGATCGATTTCACTCCACGGGTCAATCGAGCCGCTGCAAGGACGGCAACTGGGGAGCGATG
GATTTACCGGGACGTGGATCGAAGGTACCGTGCCTCGATCCTGTTGCCCGCCCGAACACATCGACGACGAT
TGGATCGGTTCTGTCGGGGTCAACGAAGCCTTCGTGCGTCCCGAAGACGAACCGATTGCGGTACAAGTCTTCTTC
GAAGGCGAGCCAATGGTCCGCTGGACGCTCTTACCCGTTACGGCATCACGGTCTGCAAAGTGCTGCTGCCCGCG

CGCCTGCTGGTGGGAAGGCCACCTGCCGCATCGAGCTCAACGTGGAGAATCCGTCGTCGCGCAGCTGGTGGCG
GACGCGATGGGCCAGAAGACCGTGGGCCAGGACCCGCGCGAACTGGGCATCAAGGTCCAGCAGATCACGTTCACT
AATAAGGATCCCTGGCGGTATTGCTGGGAGAGACCTGGATTTACGGAGCGCGGCCGCGGCATCGAACACGCG
GACGAATGCTGGACGAGCCCCGACGAGTTCGGCGCGTGGACGCTGGGGTCAGAGGCAGGCATTACCTCTGGCCG
CGCGAGACCACAACCGTTCCGGTAGCCGCGACATTCACGATCAACGACGCGGTGATCGACCCCGACACCCCGAT
ATGGAAGTGAGCGTGACGGTCAACGGAGAACGCGCCGCGGTGTGGACCGACTGGCACCTGCGCAACACCGACGCC
CGCACGGTTCTTCTGCCCCGCGAATGCTTCCGCACGCTGGAGCCAATCCGCATCGCGCTCCACGTCAAGAATCCG
CGCACGTCGTTTGGCTGGGATGGTCCACCTGGGACAAGCGCCCGGTGGAATTCGTCTCAACAACCTTCGCATT
GCGCCAATCCTGCAATACCATTTGGGCGACGTCATGGATTTTACGTCCGGTGGAGAATCCGTTCCGTTTCGACGGC
AATTCCATCGGGGTTGAATGGTCCGCTCCCGACGCTGGGGATTCTGGACGATCGGCCATAGCGCGTCCATCGTG
GTGCCCCGTGATCAACCGCTAACCGAGACGGTCCCATGGCACTGGTGATCTCCGATTGCATGGTGAACGCTGCC
GCGGCGAAGCTACCGGTCAGCGTCAAAGCCAATGGCCGACGGTCGCGGAATGGGTCTTGATAACCGCAAGCCG
CACACGCGTTCTTTTATGCTGCCCCGCGAGGCGGTAGCCGCCGCGCCCAAGCTCACGCTCACCTTCGAAATCCCC
GACCCGCGCTCGCCGGTCTCTTTCGGCTGGGGTCCCGATCCGAATCCACTGGGCTTGGCGATCGCGCGCGCCGTG
ATCGGCAAGAGCCAGATCGAGATTCCGGACTTCGAAAAACATATCCGCTACCGCACCTTGAAACGCATTCCGGGA
GTGCCGCAACTGGCCGCGCTCGTGGTCCGCATGTTTGCAGGTATCGGTAGATGCCCGGCCGCTCACCGGATTG
CGGATTGTGGTGACCATCCCGCCCCACGTCTGGTTCGGCGGCGTGGATTACGATTTTCGGGTGGAGATGTCAGAG
GAGTTGCGCACCCCTTGCGCGGAGGTCTTCGATCTGGAGATCGCCGGCTTCACAGTGGCAACGCCAACTACATC
GCAGGCGCGATTGAAGGCCTGAAGGCATTCCGGCCCGATGTGCGCTCCCCCTGCCAACGCGTTGTACGCCCTG
CTCTGCGTGACGCGGAGGGCAGGAACATTCTTACCGATGTTCTCCAAATCCCCACGGTGATGCTGTGGGATCAC
GGTTTGTGCAACTGCCGCGACAGATTCTCAAGCACCAGCCCTCCATCCCTGCCGAAGCGCAGGGGGATGCCGTG
CGGCGCTGCGCAAGGTGATGAATCACGCGCTCTACTGCCACTACTCGCCCGACCGCGGGCACGTTGCCGCTCTC
GACCAACTCGGTGTGGTAGAGGCGTCCAAGGTGCACTTCTTCTGCAAGCTGCTTATCCAACTTCGTTTCGCCAC
GGCTACCGGCCGCGGGCTCGAACGCGTTTCGCACGCGGGTCGTTTTCGCCGAAACGTCTACCTGCAAGCATCC
AAAGATCTGCCTTTTCGCAATCAGCCCGTGCTCGCGGGCATTGAGTCGCGGTCTTGCCCGCAAGAAGGCCCGC
CTCACGAATGCCTTTGGGACCTCTTCTGACGAGATCCAGGCGCTCGATAAACCCGCCCCGAAGGAACCTCCGG
CTCGAGCCCCGATTCCACCTTCTTCTGGAATTTCTGTGCATGAGGAGATCGAGTTGGTGGGCAATACCGACGTGCGC
CTGGCGGTCTCACAGGACTCAAGCGCGAATTCGAATTTACGGCAACTTCGTGAGCCTAAGAGCACCGAGACG
CTGCTCGGCCAGTACCGTATTAGTTCCGCAAGTGCTGGACTACTTCACCGAGTTGCCGCTGCTCTTTATGAAC
AGCGACGTAATCGTGGACGTCATCAATCTCGGCTACAACAGCGGCATCTCCCCGAAAGTAATGGGCTGCTTCGCC
TGCGGCGGGTTCGTTCTCTTCGATTACAAAGACGATTTCCGCAAGACCATAGGCGACGTTGCCAACAGGTGATG
TACCGCAGCGTCGATCACCTCAACACGCTCGTGGACGAGTATCTGACCGACGCGCGCAAACGGCGCGACGTCTCG
CGGTACCTGCAGCACCGGATGCCACGGAGTTAGTTTCGCGGCGCTGGCCAAACACATTCTGGTGGAC

750814 751178 365 757991 758355 365 6812 72.2826 1.6304 214 42.3288 57.6712
39.4737 60.5263 0.0000 1509169 1509169

CAGTTCGCTCCTACGTTTTCGGGGAGGAAAAAACTACGCGCGCAGGTAAGATCACTTAGTGCGCTGGCTACAG
CGCCGGTCCGAATTCGAGATTCAGCATCAAGGTGGGCCCGAACGCGTGAAGATTGGCTTCGGGTAGCATCGTAT
CTGGCGCAGCGTCACGTTTGGCTGGAACAGGATGGACAGGCGCTTCCGGCACGGGAAGTAGTGAGAAGAGTTTT
TGGGAAGGTGGGAAGTGTTGGTTACCAATACGGTTACTTTGACGGCCGGCAGGAATCTCTTTGCCATCGTCACT
GACGGCGATTCCGTTGACGTTACGCCAGGTAGGCCCGTATTCTCTTCTGGATGGCGGGATTGA
TCAATTCCACCCACAAGGAGCAGGAAGACCGCCTTCCCGAGTCACATCGACAGCGCTCCCGTCAGTCACGATG
GTGAATCGATTCTCCCCGGTATGTAGCGTAACGGTATTGGCAACCGACACATCCCCCATCCCAAAGCTCTTC
CGCACCGATTGCCGACCGGAAGCGCTGGCCATCGTGCTCCAGCCAGACTTGACGCTCCGCTGATATGTAGCG
ACACGGAAGGAGACCTTCACGGGACGCGGCGCCCCCTCATGCTGAATCTTAAATTCGGTTGTTGTTGGAGCCAG
CGCACCGAATATTCCAGGCCGCGCTGCGTGGTTGATTCTGCTCCGGAACAGTCGGTGCCATCTG

753434 755520 2087 760843 762941 2099 5322 66.5885 3.5664 550 38.3421

61.6579 37.1388 62.8612 0.0000 1516363 1516372

TGCAGGGCCGGCTGCACGAACGCTATTTCTCTCGGAATTCAACACCCAGACCAGCTACACGCCGGTTCGCTGATC
TCACCGAAGAGATCCGGCGTGACTTCTCGGTTTCGGAGCCCCGGGGCCAGCGGCTACGGCAAGTGGCCATACGTTG
TGCTTGAGCACGCCGAACCTCCGCTGACGCTGCCTGGCGGAAAAGCGTGGCCCAAAATCTCGGTGATCACGCCGA
GCTTTAACCAGGGCAAGTACATCGAACAAACGATTCTCTCCGTCTGCACCAGGGGTATCCGAATGTGGAGCACA
TTATCGTGGACGGCGCGTCGACGGACGGGACATCGCAGGTCCTGGACCGCTACAGGGATAAGGTCGCGCTGATAA
CGAGTGAGCCCCGACCGTGACAAAGCGACGCCATTAATAAGGGCATGGCGAGAGCCAGCGGTGAGATTCTGACGT
GGATAAACAGCGACGATATGCTCGCCCCGGGTGCGCTCGCAGCCATGGCTATGGCCTTCGAGACTTCCGGCGCCG
ACATGGTGGCCGGGATCTGCCAGCTATACCGGGACGGGCGTCTCGTGAATCAGCACCTCACCTCGTGTCCGGACG
GTCCCCTGAATCTTAACGATCTGCTCGACCTGGACGATTGCTGGAACGAAGGGCAGTTCTTTTACCAGCCGGAGG
TAATGTTACACAGGAACTCTGGGTGCGATCGGGAGGGCACGTCAAAGAGACCCTCCACTACAGCATGGATTACG
AGATGTGGTTGCGCTTTGCCAAGGCCGGCGCAACCTCCACGTAATCGGGCGCCCGATCGCGTGGTTCAGGGTGC
ACGACGAACAGAAAAATATCTACCATCAATTTTCTACCCGAGTTGGTTGCATGCCGCGATGCGTTTGCGAAGG
AACATAGGATTGCCCTCAGGCCCTCGTTAAATGTTCTGGCAAAGCCATCCAAGCTGCGCATCACGCTGCTCAACG
ACCACGGCGGGTCTATGGCGCGGGCATCGCACACGTCCGGCTTGGGAGAGCCTTGGCCAGGGCAGGGCACGATG
TGAATCTGGTGTGATCTTAGACAACCCCAACTCTGGGGCCGAGTTTGGGGATTACACCACCCAGAGCGTGGTCA
ACCGTGTGGGCGCCTCCAACCCGGACCTGGTCATTGTGCGCAATCTGCATAACCGGAAGGTCGACCCGATGCTGT
TGACCTGTTGAGCGAGCGCTACCTGACGCTCGTGGTGTGACGATTTTGGACCCTGACCGGCCGCTGCGGCT
ATACGGGCGATTGTGAGAAGTATTTAACCGGATGTGACGCGACGTGCCCAACTCCCGACGAGTATCCGGCCCTGG
CTCCCGCCGAAATCGCAGAGTCTGGTCAAAAAAGCGGGCAATCCTGGGGGAGCCCCGTCGTCCGGTCTCTGTTGG
CGAACTCGCAGTGGACTGCCGAATTCGCCCGCAACGCATTTGCCAATGGTTCGGGCGGAGACGTTCCATCCATTG
ACATCTGCCAACTTAGCTTCCCGCTCGATATTTTTCGACCTCGGGACAGACGCGCCTGCCGCGAGACTCTCGGGC
TGCCCCTGGATCGTTTCATCATCCTGCTTCCGGCCTCGCTGGAGGACCAGCGAAAAGGGAGGCACCCCCCTCCTGA
ACGCCTTATCCCGCCTCGAGCTACCCGATGTCCTGGTGGTAACTCTCGGGGCCGTAGCCCAAGGCGTGGAGGCCG
GAATCGACCTGGTCCAGCTAGGCTTTATTTACGATCAGCCGAGAGTGGCGATGCTTTATTCCGAGCCGACCTCG
TGGTCCGGCCCCAGCACCTCCGAGACTTTCCGGCAAGTCTTCATTGAGGCAATTGCGTGCGGCACCCCCGGTTCGAG
GTTACGCGGTTGACGGTGTACGTGAGGCAATCCGTGACGGCATTACGGGCGTTCTTGCGGCGGACATCGATCCGT
CCTGCCTGGCATCCGCGGTTCAATATCTTTACAATCGTCCTGACCTGCGAAGGGACATGACACACTGGGGACGCT
CGTTCGTAGAGAACGAGTGGTCCGAGTTTTTCAGCGTACCGGCACCTGTTCTGGCGCTGCGG
CCGAAGCGACAGCAGAAGATGCCGGGACGCCGCAATTCCGACCACTCGTTTTCCACGAAGAGACGCCCCCATCG
AGAGAGATCCCGGCGAATCCCGGGATGAGCGTACAGATATTGAACCGCTGCCGCGAGGCTCGCCGGATCGTCTCC
CTGAGAGAGTAGGCCGGTGACGCCGTGCGGGATCGCTCCGCGGATTCCGTTGATCGGATAGCCGAGACCGGTGT
TCCGCAAGCGATGGCCTCGATAAAGGTCTGCCCGAAGGTCTCTCCGTGCGCGGCGCAACTACCAGATCGGCCGC
GGAGTTAATCATCGCCACTTGGTCTGGCCGTTGATCTGCCCGAGCTGACGGACTTCGACTGGGAAGGTGCGGCC
CGGGTCGGGCGCGCCCATCGTGGCCACTAACAGTTTCGGCAACTCCAGGCGCCCCAGCGCTTCGAGGAAAAGCCCT
GCCTCCTTTTCGCGGATCTTCCACGGCTCCCGCGAGGAGAATCACGAAGCGATCATAGGGCAGATCGAAATTGTC
GCGGCACATGCGCCGGTTCGCGGAGCGGAAAATGTGGAGTGGAAAACTCAGTTGGAAGCGCGCGATGGGCGGAGT
GCGGGGAAGCTCGTGAGCGTTGGCCGAAAAGCATTGCGGGCGAGCTCAGCCGCCAGCGGGTACTGCACAGGAT
CGCGGGCCCTTGCGCCTCGCCGAAGATGAGGCGCTTGTGGACCAGGCATCGGCGATTTGCTCCGGCGGCGAGCGC
TGGGTACTCGTTCCGGCGTGGGACATGTGTGATCGCAGCCTGCCAGGTATCGATCGCAGCCAGCGGGAAAACGCGCA
TCTGCCGGTCAACGCCCAGTAATCGTGAGTACGACCGCGGACGGAAATCGCTCGCAGAGCAGTTGCAGCAGGAA
CGGATCAGCGTTGGCGGCGTGAGATTCCCGATGATCACCAGATCCGGGTTGGTCTCCGCCACCCGATCGATCAC
CGCCTGCGAAGTGAATGGCCCGTCGATTCGAAACCGATAAGCGAGCGATCCAGAATTGAGACCAGGGTGATCTC
GTGGCCGGCCAGGCCAGGGCGCGGCCATGCGGACATGCGCAATGCCCGCGCCAAAGAAAGGGCCGATGTCGTT
GAGCAGCGTGACGCGTAGCTTTTGGCGCAGCGGAACGGCGGGTACGGCGGAGGCTCCTCGCGGCCGCTGCGTTT
CAGGAAATCGCGGCGACAGACCAGCAGTTCGGCTTGAAAGCGGGCCGTGATGTGCGTTTTCTGTTGGTCATGCAG

CCGGAACACGCGACCGGACGGCCGATCACGTGAAGCTGGGCACCGGCTTCCGCGAATCGCAGCCACATTTCTGTA
GTCCATGCTGTAGTGAGCCAATCGTTGACGTAGCCGCCCCGCGGAAGCCATAACTCACGGGTAAACATCACCTC
GGGTTGGTAGAAGAACTGGCCTGCATTCCACCCGTGGTCGAGATCGAGCAGATCTTCGAGCGGCAATGGGCCTTC
CGCGCAGGATGTCATGTGCTGGGCTTCGAGGCGCCCGTCACGATACAGCCGGCAAATGCCCCGCGATCATATCGGC
GGAGTTGGTGTCGAACGCCAGCGCCACCGCGGCGAGCGCGCCGGGGGCCAGCATATCGTCGCTATTACGCCACGT
CAGGATTTCCGAGTGGCGAGCGCCATCCCTTTGTTGATGGCGTGGCTCTGTCCGTTGTCCGGTTCGCTGATGGT
GTGCGCCAGGCGGTGCGGGTAACGATCCAAAATCGCGGAGGTGCCATCGGTGGAAGCTCCGTCCACCACGATGTG
CTCCACGCTTGATAGTTCTGGTGGAGTACGGAGAGCAGAGTCTCTTCGATATAACCGCCCTGGTTGAACGACGG
CGTCACGATGGAGATCCTGGGCCACGGTTTCCCGCTCGGCAGATTGGCGGCAGCTCGATGGAATCAAGCGGTGG
GAGGTAGGGCCATGACGCGTAGCCGCGGCGGTGGGATTGAGCCTCGGAATTCGCGGCGGATGTCTACCGTGAC
GTCGGCAACGGGAGTGTGGTTCGCGGGCGTCCAGGATTTCGCAAGAAAGTAGCGCTCGTGCAGGCTACCCTGCA
769699 769735 37 770490 770525 36 754 86.4865 2.7027 47 47.9452 52.0548
46.8750 53.1250 0.0000 1540225 1540225
AACTGGCCATAGAGGGTCACTTCAAAACGGGCCATAG CTATGGCCAGTTTTGAAGTGACCCCCGAGGCCAATT
789230 790990 1761 800924 802681 1758 9933 78.0776 2.1922 1530 38.9031
61.0969 38.5169 61.4831 0.0000 1591914 1591917
CGTCGTTACCGCGGACAGCGTGACGCCCCCGCAGGGTCGGGATCGAGCCAGACGTTCCGCTTGAATACTCGGA
CACGGCGGGCGGGCTAACCTGTCCACGGCGTGGGGATGGTTCAACGCCAGCTTTGCGTCCAGTCCCGCCAATTC
CTGCATGATCTACTACGATCGGGCGGCGGCACGTACTATTTGCTCAACGATGCCGGGACTCAGTGGATCCAGGG
CGCTCCGGGGGTAAACAACGACTATCCAGAACAGCCAGTGCTCTGTGAATCTGTCCGCCAGCAGTGCCGTAATCAG
CGGCAACGTCTCACCCTCAATCTCGCCATGACGTTCAAGGCAGGGTATGCGGGAGCGAAGAACGTCTATATGTA
TGCCGTCGATCCGACCGGAGCGAATAGTTCTGTGGCAGACGCGCGGAACGTGGACGGCACAGGGCGCGATTGTGAC
CGCGGACAGTGTAACGCCGAGTTCCGGGTCCGGGACGAGCCAGGCGTTTCGATTACAATACTCGGACACGGCGGG
CGCGACCGGTCTCTCGACGGTGTGGGCGTGGTTCAGCCCCGACCTTGCCGCCAGTTCCGGCGAGTTCTGTGCATGGT
CTATTACGATCGCGCGGCGGGGACTCTCTACCTGATTAACGATGCGGGAACTCAGTGGCAACTGGGCTATCCGGG
AACAACGGCGACGATGCAAAACAGCCAGTGTTCCATCAATCTGGCCACCAGCGGCGTGGCTCTCAGTGGCAGTAC
GTTAACGTTGAATTTGGCGATGAACCTCAAGCCGACGTTTGAGGCGCGAAAAATGTCTATATGTACGGAGCCGA
CTTCGGCGGTTCAAACAGCGGATGGCAGACGCGTGGGACCTGGACTGGGACAAGTGGAGCGGCTACAGTGACTGC
GGACAGCGTGGCTCCGAGTTCGGGGTCGGGGTCGAGCCAGACGTTCCGCTTGAGTATTCCGACACGGCTGGAGC
GACCAATCTCTCGACCGGTGGGTCTGGTTCAGCGCGACGTTCCGGCAGTGCGGCGAACTCCTGCATGGTCTACTA
CGATCGTTTTGGCGAGTACTCTCTACCTGATCGACGATGCGGGAACTCAGTGGCAGCCGGGAACTCCGGGGATAGC
GGCGACGCTGCAAAACAGTCAGTGCTCCATTAACCTGACAGGCACAAGTGTGGCTCTCGCCAGCAATACGTTGAC
GCTGAATCTTGCGTTGACGTTCAAGCCCGCTTACGCCGAGCGAAGAAGCTTTACATGTACGGGGCCGATGCCTC
GACAAATAGCGGTGGCAGACCCGTGGGACGTGGACCGCAACCGCCGGAGCCGCTGCCGTGACAGCGGACAGCGT
GACGCCGAGTTCGGGGTCGGGCTCGACCCAGACGTTTGTATTACAGTATTCCGACACCGCCGGGGTGACCACTCT
CTCGACCGTGTGGGTCTGGTTCACGCGACGTTCCGGCAGCGCAGCGAACTCCTGCATGATCTATTACGATCGCGC
GGCGACCACTCTCTACCTGATCAACGATGCTGGAACCTCAGTGGCTGCCGGGCGCTTTGGGATCCGGGACGATTGA
AAACGGCCAGTGTTCCGTCAACCTGACAAGCAGCAGCGTTCTGCTGAACGGCAACGTTCTAACCTCACTCTTGC
CATGACCTTCAAACCTGCCTATGAGGAGCGAAGACCATCTACATGTACGGGGCCCGCGGCCACCAATAGCGG
GTGGCAGACACGCGGAGGTTGGAGCGCACAGTAGCA
TGCTATTGGGCTGTCCAGGTCCCGCGCGTCTGCCATCCGCTGTTAGTCGAGGCATCGGCCCCGTACATATAGATG
TTCTTCGCTCCGGCGTAGGCGGGCTTGAATGTCAACGCGAGGTTCAAGTGTGAGCGTATTGCTGGCGAGAGCCACA
CTTGTCCTGTGAGGTTAATGGAACACTGGCTGTTTTGAGCGTCCCGCCGCTCCCGGCGTCCCGGCAGCCAC
TGAGTTCCGGCATCGTTGATCAGGTAGAGAGTGGTCGCCGCGCGATCGTAATAGATCATGAGGAGTTCGCCGCA
CTGCCGAACGTGCGCTGAACAGGCCCATGCGGTGAGAACTGGAGGCCCCGGCGGTGTCCGAATATTGTAAC
GCAAACGTCTGGCTCGAGCCTGATCCCGAACCCGGTGTCACTGTCCGCTGTACCACAGCGGCTCCGGCGGT

GCGGTCCAGGTCCCACGGGTCTGCCATGCGCTGTTGGTCGCTCCCGCGGCGCCGTACATGTAGACGTTCTTCGCT
CCGGCGTAGGCTGGCTTGAATGTCAACGCGAGATTAGTGTACGCGTATTGCCGGCGAGAGCCACACTCGTGCCCT
GCCAAGTTAATGGAACACTGGCTGTTTTGCAGCGTCGCCGCCGCTCCCGCGGTGCCCGGAGCCACTGAGTTCCG
GCATCGTTGATCAGGTAGAGAGTGGTCGCCGCGGATCGTAATAGATCATGCAGGAGTTCGCCGCACTGCCGAAC
GTCGCGCTGAACCACACCCACGCGGTGAGAGACTGGACGCCCCCTAGCGAATCGCTGTACTGCAACGCAAACGTC
TGGCTTGAGCCTGATCCCGAACCCGGTGTACGCTGTCTGCTGTACGACAGCGGCTCCGGCGGTTGCCGTCCAG
GTCCCGCGGGTCTGCCATCCGCTGTTGGTTGCGCCCCGCGGCGTCGTACATGTAGACGTTCTTCGCTCCGGCGTAG
GCTGGCTTGAATGTCAACGCGAGATTAGTGTACGGGTATTGCCGGCGAGAGCCACATTTGTGCTGCGAGGTTA
ATGGAACACTGGCTGTTTTGCAGCGTGGCCGCCGCTCCCGAGTACCCGGTAGCCACTGAGCTCCTGCATCGTTG
ATCAGGTAGAGAGTGGTCGCCGCGGATCGTAATAGACCATGCAGGAGTTCGCCGCACTGCCGAACGTCGCGCTG
AACCAGACCCATGCGGTGAGAGACTGGACGCCCCGGCGGTGTCCGAATACTGTAATGCAAACGTCCTGGGTGAA
CCCGACCCCTGAACTCGGCGCCACGCTGTCCGCTGTACGGCAGCGGACCCGGCGGTTGCGGTCCACGTCCACGG
GTCTGCCAGCCGCTATTTGTGAGGCATCGGCCCGTACATGTAGATGTTCTTCGCTCCGGCGTAGGCGGGCTTG
AATGTCAACGCGAGATTAGTGTCAACGTATTGCCGCCGAGAGCCACACTCGTGCTGCCAGGTTAATGGAACAC
TGGCTGTTTTGCAGCGTCGCCGCTATCCCCGAGTTCGCCGCTGCCACTGAGTTCGCCGATCGTTGATCAGGTTT
AGGTTGTTGGTCGAGCGATCGTAATAGACCATGCAGGAGTTCGCCGCACTGCCAAACGTCGCGCTGAACCACACC
CATAACGTTGAGAGACTGCTCGCTCCAAGCGTTTCGCTATACTGCAACGCGAACGTCTGGCTCGACCCCGACCCC
GAATTTGGCGTCACACTATCTGCTGTACGCGG

935160 935193 34 935247 935282 36 53 88.8889 5.5556 48 50.0000 50.0000
50.0000 50.0000 0.0000 1870440 1870440 AAGGCATCTCTCGCAAAGGCGCAAAGAACGCAAA
TTTGCCTCTTTGCGCCTTTGCGAGAAGACTGCTTT

935160 935282 123 935465 935586 122 182 73.4375 8.5938 64 51.4286 48.5714
51.0638 48.9362 0.0000 1870747 1870745
AAGGCATCTCTCGCAAAGGCGCAAAGAACGCAAAGAAACCAACACAGAATCAGCCAGGGATCTGCTGCTTTCTT
CAGGTTCTTCTTTTTGCGTCCTTTGCGCCTTTGCGAGAAGACTGCTTT
AAGGCATCTCTCGCAAAGGCGCAAAGAACGCAAAGAAACCAACACAGATTAGCCAGGGATATCAGCTGCTTTC
TTGAGTTCTTCTTTGCGTCCTTTGCGCCTTTGCGAGAAGACTGCTTT

935465 935501 37 935548 935586 39 46 89.7436 5.1282 54 51.3158 48.6842
51.4286 48.5714 0.0000 1871049 1871049
AAGGCATCTCTCGCAAAGGCGCAAAGAACGCAAAGAA
TTCTTTGCGTCCTTTGCGCCTTTGCGAGAAGACTGCTTT

965933 965986 54 966006 966060 55 19 78.5714 5.3571 46 41.2844 58.7156
40.9091 59.0909 0.0000 1931992 1931991
CCGTAAACGAGCGCACCAATCCGAGCCGCGACCGTTAAGGGAGCGGTGCTAAT
ATTACCGACAGCTCCTTAACAGTCGCGGCTCCGATCATTGAGCGCTTGTCACCGG
972472 972494 23 972537 972559 23 42 95.6522 0.0000 41 32.6087 67.3913
31.8182 68.1818 0.0000 1945031 1945031 CCCGTCACGCAAGTCGCGTGACG
CGTCACGCAACTTGCGTGACGGG

1014409 1014437 29 1014526 1014554 29 88 89.6552 0.0000 43 36.2069 63.7931
34.6154 65.3846 0.0000 2028963 2028963 TCCGAGCCGCGACCGTAAGGAAGCGATCT
AGATCGCTCCCTTGCGGTGCGGGCTCTGA

1081082 1081134 53 1081157 1081208 52 22 83.0189 1.8868 59 51.4286 48.5714
52.2727 47.7273 0.0000 2162291 2162291

GATTCTAGAAAAGGCAGTTTCTCGCAGAGGACGCAGAGATCGCAAAGAAAACC
GGTTTTCTCTGCGATCTCCGCGTCTCTGCGAGACATAATTTTTCCAGACTC

1120857 1120886 30 1120907 1120938 32 20 90.6250 6.2500 45 46.7742 53.2258

44.8276 55.1724 0.0000 2241793 2241793 CTTTGGCGCTTTGGCGCTTTGCGAGAAAGG
CCTTTCTCGCAAAGCCGCAAAGTAACGCAAAG
1120845 1120871 27 1120914 1120940 27 42 100.0000 0.0000 54 51.8519 48.1481
51.8519 48.1481 0.0000 2241785 2241785 TTCTTTGCGTTACTTTGGCGCTTTGCG
CGCAAAGCCGCAAAGTAACGCAAAGAA
1259344 1259378 35 1259384 1259419 36 5 88.8889 2.7778 50 33.8028 66.1972
31.2500 68.7500 0.0000 2518762 2518763 CAACGAGCCGCGACCGTTAGGGAGCGCTACCCCTG
CAGGGCGTAATGCTCCCTTACGGTCGCGGCTCGTTG
1278293 1278322 30 1278324 1278353 30 1 96.6667 0.0000 55 30.0000 70.0000
27.5862 72.4138 0.0000 2556646 2556646 AACGAGCCGCGACCGTGAGGGAGCGCTACC
GGAAGCGCTCCCTCACGGTCGCGGCTCGTT
1378025 1378056 32 1378083 1378114 32 26 93.7500 0.0000 54 46.8750 53.1250
46.6667 53.3333 0.0000 2756139 2756139 TCTCGCAGAGGACGAGAGATCGCAAAGAAAA
TTTTCTCTGCGATCTCCGCGTCTCTGCGAGA
1398749 1398788 40 1398807 1398846 40 18 95.0000 0.0000 70 45.0000 55.0000
44.7368 55.2632 0.0000 2797595 2797595
TTTTCTCGCAGAGCCGAGAGCCGAGAGTAACGCAAAGA
TCTCTGCGTTACTCTGCGGCTCTGCGTCTCTGCGAGAAAA
1427873 1427918 46 1428547 1428593 47 628 87.2340 2.1277 62 51.6129 48.3871
53.6585 46.3415 0.0000 2856465 2856466
TTTCTCGCAGAGACGAGAGGCGCAGAGGAACGCAAAGAAGATAAA
TTTCTCTCTTTGCGTTACTTTGCGGCTCTGCGTCTTTGCGAGAAA
1511665 1511754 90 1512087 1512178 92 332 78.4946 4.3011 78 42.3077 57.6923
43.8356 56.1644 0.0000 3023841 3023841
TTCTTTTCGCGTTGCTGGGAGGTTTGCGGCGTTGCCAGAGGTCGGGAATTTCTCGCAGAGCCGAGAGCCGAGAG
GTCACGCAAAGAAGA
TCTTTCTCGGTGACTCTGCGGCTCTGCGGCTCTGCGAGAGACTGAGAGTTTCTCGGAAGCCGCCAAGCCGCAA
AGTAACGCCAAGAAGAA
1512087 1512130 44 1512134 1512177 44 3 81.8182 0.0000 48 42.0455 57.9545
41.6667 58.3333 0.0000 3024264 3024264
TCTTTCTCGGTGACTCTGCGGCTCTGCGGCTCTGCGAGAGACT
AGTTTCTCGGAAGCCGCCAAGCCGCAAAGTAACGCCAAGAAGA
1522520 1522554 35 1522571 1522605 35 16 88.5714 0.0000 50 51.4286 48.5714
51.6129 48.3871 0.0000 3045125 3045125 ATTGGCCGAGATGAACGCCGATGAACACAGATAA
TTATCTGCGTTTATCTGCGTTTATCTGCGGCTCTGCGGCTCTGCGAGAGACT
1582455 1582490 36 1582515 1582550 36 24 83.3333 0.0000 42 56.9444 43.0556
60.0000 40.0000 0.0000 3165005 3165005 TTTGGCCACGAATGCACACGAATGTCCACGAATAAA
TTTATTGCGGTATATTGCTGTCCATTGCTGACCCAA
1629056 1629111 56 1629138 1629190 53 26 78.5714 5.3571 46 39.4495 60.5505
36.3636 63.6364 0.0000 3258249 3258249
GGCCCCCTTTCCCTCATCTCAGCTCTTTCTTCAGAACGGTGGGGATCGGTCAAGGG
CCCTTGACCGATTCCCTCCGTTCTGGAATCTGGTACCCGAGGGAAAGGGGGCC
1922697 1922737 41 1922755 1922793 39 17 85.3659 4.8780 48 45.0000 55.0000
42.8571 57.1429 0.0000 3845492 3845492
ATTCCTCGCAGAGACGAGAGCCGAGAGACCCGCAAAGAA
TTCTTTGCGTTACTCTGCGGCTCTGCGTCTGCGAGAAAT

2114413 2114462 50 2114582 2114631 50 119 84.6154 7.6923 56 35.0000 65.0000
36.3636 63.6364 0.0000 4229044 4229045
TGC GTATTCCGGGGCAAGCCGAACAGCATTCCGGGGCTGGCCGAACAGGC
GCCTGTTCGGGCGCATCCCGGAATGCTGTTTCGACTTCGCCGGAATGCGCA
2359212 2359250 39 2359267 2359305 39 16 87.1795 0.0000 53 48.7179 51.2821
47.0588 52.9412 0.0000 4718517 4718517
TTTTCTCGCAGAGGCCGAGAGATCGCAGAGAAAAACCAA
TTGGTTCTCTCTGCGTCCTTTGCGGTCTCTGCGAGAAAA
2383422 2383454 33 2383477 2383509 33 22 87.8788 0.0000 46 42.4242 57.5758
41.3793 58.6207 0.0000 4766931 4766931 CTCGAGAGGACGCGGAGATCGCAGAGAAAAACC
GGTTTTCTTTGCGAGCTCTGCGTTCTCTGCGAG
2416753 2417269 517 2426192 2426711 520 8922 73.4733 2.0992 331 38.5728
61.4272 38.7013 61.2987 0.0000 4843461 4843464
GTTGGCGAAGAAATTAGTCGCGTTGAGCACTTTGTTACGGAAGTACTCGTACGCCGACCCATGAAAATCGTTGGT
GCCGCTGCGCGAAGCCATATTCATGACGCCGCCGGTATAGCGGCCATACTCAGCCGTGTTGTTGCTGGTCTGAAT
GCGGAACCTCGGAGACCGCGTCTGCGCTCGGTGTGAGGGCAACGATGTTGCCGTAGGTGTCATTACCGGCACGCC
GTCGTAATACGTTGCGCTCTGGTTGCGCGTGCCGCCGCGCATCTGGTAATTGCCCCGCGGCGAAGACGTTCTTGCC
GGTGAGGCTCCCTTCAGAGCTGCCCTGCGGCACCAACCCCGCGCGAGCGAGATCAAGTTCAGGATATTGCGCCC
GTTTCAGCGGCAGTTCTTGACCGCGCGGTGCCGACCACTGGCTCAAATTGGCCGTTTCCGTTTGACGAGCGCG
CGCTTCGGCCTGCACTTCCACGGAAGTGGTTGACGTACCGCACTTGCAATTGCGATATCCGCGCGCACCC
GGTCCGCGCGGACATTTGTTGAGCTCGGCGACGTACCCAGACAGTGGAAGTCCAGGCGGCGGCGCCACTGCT
GAGGACCGAAGACGCCAACCTGAGCCAGGTGCTGAGCTCCAGAGCCGTAGAGGACTTGCCGGTGAACGGCCGCAA
CATTCTCAACCTCACGGCACTGGTTCTGGCGTGGTGCCGCAAGGCACGACCGATGGAAACGCGATCACCGGCAA
GAATATCTTCGCCGCCGGAATTATCAGATCGGCGGCGGCGCGGCAACAGGGCGCGGTCTATTACGACGGCGT
TCCCGCAATTCCGCGTTGGGAAATCTGGTGAACATGGTGCCGAGCCCGGACGCGATCGCGGAGTTCCGAGTTCA
AACCAACAGCAACAACGCCGAATTCGGGCGTTACGCGGGCGGCGTGGTGAATATCTCTTCGCGATCGGGCAGCAA
CCAGTATCATGGCAGCGGTACGAGTACTTCCGCAACGATGTCTGAACGCTACGAATTCTTCGCCAAC
2422870 2423579 710 2431595 2432304 710 8015 69.8324 1.6760 328 41.0563
58.9437 42.8000 57.2000 0.0000 4855174 4855174
CCCTGCTTTTCGAGATAGGCCTTCGCGATCCTCGCGTTTTCCAGGCCCCAAGCCGGTTGGGGCGGTGGCGAATCG
AGCTCCACCATTACCAGCCCCGCGTCTGGCGGCCTTCCATCATCGAAAGAATCGCCGGAACATCGACCTTCCCC
TGGCCGAGCGGACAATATCCGAGATACTCCTGGCCGCTGAATAATCCTTCAGGTGCATATGCTGCACCAGCGGA
AGAAAATGCTTCACCACCTGCACCGGGTCCGATCCGCCCTTCTGCAACTGACCGATGTCCGGCCCCGAACCTTCATT
ACTTTCTGATCCACAGACTCCATCACGGCGTACGTCTCATCGCGCGTTTCGACGCGAGTTCCCGTATGCTGATGC
AGAACCGGAGTCAGGCCCATATCGGTACGCGCCTTTCCAAGTTCGTTGAGCGTGGTGAGAATATCGTCCTTATGC
GCTTTGAAATCGAAGCTGTGCGCGCCGACACCATTCGGTCCCACCACGATCACTTTGCCGTTGTACTTCCTAACG
AGCCTCGCCAGTCCAGGGTTTTTGCGATCGAATCCTTCGGGCGCGGGATCGGACAGATTGGTCGAGCAGTAT
GCCGAGATGAGCGGCGAGTTATGCTGTGCGAGCAGGGCGCGGATGCCACCCTCCGCTCCATCGCTTCGATCTGG
TTGCCAAACAACCTCGACGCCGTAAAAGCCGAGCCC
GGACTTGCTATTACGGGCTCGAAGCTTTCAGCTTCCAGGTTTCAGGGAATGGAGGCGAACGGCGGGATCGCATCT
CTGATCGAAAAATACAACCTCCCGTTGATTTCCGCGTACGGCGGGCCGAACCTCATGGATGCCGGGCGAGCGTCAG
GAAAACATCGCACAGGCGATCGAGTGGGGAAGCTGCTGAAGAAGTACGGCGGCAAAGTCGTAACCATCGGCCCC
AATGGCGTGCGGCGCGATTCTACAATTTCAAAGAGCATAAGGACGAAATCGTCACCACGCTGAATGAATACTGC
AAAGCGCTGACCGATATCGGGCTGACCGGATGCCTTCATCAGCACACCGGCACCTGCATCGAATCCCGCGACGAG
ACGTACGCCGTGCTGGAAGCGGTAGACACGCGGTACGTCAAGTTCGGTCCGGATATCGGGCAGCTTCAGAAAGGC
GGCTCCGACCCCGTGAGGTGGTGAAGGACTTCTGCGCTGGTGAATCACATGCACGTGAAGGACTACACCGGC

GGAGAGGCGTTCGTGGGCTACTGCCCCGCTGGGACAAGGCAAGGTCAACATTCCCGCAATCCTCGACATGATGGAC
GGGCGGCCCCACCGCCGCGCATGCTCATGGTCGAACCTGACGCGCGGACCGCGAATGCCGATGACTGCGCTCGAGACG
GCAACGATCGCAAAGGCGTACCTGGAAAAACAGGG
2575975 2576009 35 2576022 2576056 35 12 94.2857 0.0000 60 34.2857 65.7143
33.3333 66.6667 0.0000 5152031 5152031 AGCGGACGCTCCCTAACGGTCGCGGCTCGGATATC
GACATCAGAGCCGCGACCGTTAGGGAGCGTCCGCT
2852073 2852250 178 2854204 2854380 177 1953 91.5730 0.5618 279 37.1831
62.8169 36.8098 63.1902 0.0000 5706454 5706454
GAGCCGCGACCGTGAGGGAGCGTCCGTTGTACGTGTAACACGATTCCAGTTTGCGGCCCTGCCGCTTTGTGGG
GCAGGCAATCCTGCCTGTCGCGATGGCGCTGCGCGCCATTTAAGGGAATGAAAACCCCTTTGTGGGGCAGCCAAT
CGTGGCTGCAGCCGCTTTTCAGGCGGCT
AGCCGCCTGAAAGGCGGCTGCAGCCACGATTGGCTGCCCCACAAAGGGGTTTTTCATCCCCTTAAATGGCGCGCAG
CGCCATCGCGACAGCCAGGATTGGCTGCGCCACAAGGCGGCGAGCCGCAAATTGAGGGCGTGTTTAAACCGTACA
ACGGACGCTCCCTTACGGTCGCGGCTC
2884685 2884724 40 2884752 2884791 40 27 92.5000 0.0000 65 37.5000 62.5000
35.1351 64.8649 0.0000 5769476 5769476
AGATTAGCGACCGCTCCCTGACGGTCGCGGCTCGGATGGA
TCCTTCCGAGCCGTGACCGTCAGGGAGCGGTCACTAATCT
2888071 2888124 54 2888129 2888183 55 4 81.8182 1.8182 58 52.2936 47.7064
51.1111 48.8889 0.0000 5776253 5776254
GCTCTTAGTCTTATCCGTGTTGATCCGCGTTCATCCGCGGCCAATTCTGCTTTT
AAAAGCCAATATAGGCCACGGATGAACACGGATGAACGCGGATAAGAAAAAGAGC
2898340 2898376 37 2898394 2898431 38 17 89.4737 2.6316 54 30.6667 69.3333
29.4118 70.5882 0.0000 5796770 5796771
CCGGCGAACC GCGACCGTGAGGGAGCGATCGCTTGCT
AGTAAGCCGATCGCTCCCGCACGGTCGCGGTTCTGCTGG
2904896 2904933 38 2904949 2904987 39 15 84.6154 2.5641 46 53.2468 46.7532
54.5455 45.4545 0.0000 5809882 5809882
TTTCTCGCAAAGGCGCAAAGAGCGCAAAGAAAACCAA
TTTGGTTTTCTCTGCGTTCTTTGCGGTCTCTGCGAGGAA
2904945 2904994 50 2905272 2905322 51 277 82.3529 1.9608 55 61.3861 38.6139
61.9048 38.0952 0.0000 5810266 5810267
TTTGTGTTGTTTTCTCTGCGTTCTTTGCGGTCTCTGCGAGGAATATTCTT
AAGAATTTTTCTCGAAAGAACGAGATCGAAAGAAAACCAAAGAAA
2907012 2907049 38 2907144 2907181 38 94 100.0000 0.0000 76 44.7368 55.2632
44.7368 55.2632 0.0000 5814193 5814193
TCTCGAGAGCCGCAAAGCCGCAAAGTAACGCCAAGAA
TTCTTGCGGTTACTTTGCGGCTTTGCGGCTCTGCGAGA
2919318 2919350 33 2919399 2919432 34 48 91.1765 2.9412 51 44.7761 55.2239
41.9355 58.0645 0.0000 5838749 5838749 GGCTTTCTTTGCGGCTTTGCGGCTTTGCGAGAA
TTCTCGCAAAGCCGCTAAGCCGCAAAGTAACGCC
2964326 2964358 33 2964386 2964419 34 27 97.0588 2.9412 61 37.3134 62.6866
36.3636 63.6364 0.0000 5928744 5928744 CGGAGCCGCGACCGTAAGGGAGCGGTTGGTAAT
ATTACCAACCGCTCCCTTACGGTCGCGGCTTCCG
3088962 3088996 35 3089013 3089047 35 16 91.4286 0.0000 55 31.4286 68.5714
31.2500 68.7500 0.0000 6178009 6178009 GACACCAACGAGCCGCGACCGTCAGCGAGCGTTAC

GTAGCGCTCCCTGGCGGTCGCGGCTCGTTGGTGTC
3102411 3102458 48 3102465 3102512 48 6 77.0833 0.0000 41 48.9583 51.0417
45.9459 54.0541 0.0000 6204923 6204923
ATCTCTCGCAAAGCCGCAAAGCCGCAAAGAACCGCTAAGAAGACAGAA
TTCTTGCTTCTCTGCGTTACTCTGCGGCTCTGCGTCTCTGCGAGAAAT
3110888 3110921 34 3110947 3110979 33 25 94.1176 2.9412 56 52.2388 47.7612
53.1250 46.8750 0.0000 6221868 6221868 GTTGTTCTTCTTGTCGTTACTTTGCGGCTTTGCG
CGCAGAGCCGCAAAGTAACGACAAGAAGAAAAC
3110907 3110988 82 3111010 3111091 82 21 79.5181 2.4096 77 47.5610 52.4390
48.4848 51.5152 0.0000 6221998 6221998
CTTTGCGGCTTTGCGAGAACCAGCAGTTCTCTCGCAGAGCCGAGAGCCGCAAAGTAACGACAAGAAGAAAACCA
ACAACGA
TCGTCTTTGTTCTTGTTCTTAGCGTTACTCTGCGGCTCTGCGGCTCTGCGAGAACCAACAGCTCTCTCGCAAAGC
GACAAAG
3111023 3111062 40 3111074 3111113 40 11 85.0000 0.0000 50 46.2500 53.7500
47.0588 52.9412 0.0000 6222136 6222136
TGTTCTTAGCGTTACTCTGCGGCTCTGCGGCTCTGCGAGA
TCTCGCAAAGCGACAAAGCCGCAAAGTAACGCCAAGAACA
3168680 3168713 34 3168736 3168769 34 22 88.2353 0.0000 48 41.1765 58.8235
40.0000 60.0000 0.0000 6337449 6337449 TCTCGCGGAGGACGCGGAGATCGCAGAGAAAACC
GGTTTTCTCTGCGATCTCTGCGGTCTCTGCGAGA
3276179 3276220 42 3276225 3276267 43 4 81.3953 2.3256 44 58.8235 41.1765
60.0000 40.0000 0.0000 6552445 6552446
TATCTGTGTTTCATCTGCGTTTATCTGCGGCCATATTGCTTTT
AAAAACAGTAATGGCCACCGATGAACACAGATGAACGCAGATA
3324027 3324061 35 3324077 3324111 35 15 82.8571 0.0000 40 50.0000 50.0000
44.8276 55.1724 0.0000 6648138 6648138 CGCAGAGCCGAGAGTCACGCAGAGAAAATTAAAC
GTTTTTCTTCTCAGTGTTACTCTGCGGCTCTGCG
3324012 3324054 43 3324094 3324134 41 39 83.7209 4.6512 47 53.5714 46.4286
50.0000 50.0000 0.0000 6648148 6648146
GTAAAAGTATTTTCTCGCAGAGCCGAGAGTCACGCAGAGAAA
TACTCTGCGGCTCTGCGTCTCTGCGAGAAAAAACTTTAC
3370514 3370569 56 3370596 3370648 53 26 78.5714 5.3571 46 39.4495 60.5505
36.3636 63.6364 0.0000 6741165 6741165
GGCCCCCTTTCCCTCATCTCAGCTCTTTCTTCAGAACGGTGGGGATCGGTCAAGGG
CCCTTGACCGATTCCCTCCGTTCTGGAATCTGGTACCCGAGGGAAAGGGGGCC
3388042 3388077 36 3388612 3388647 36 534 88.8889 0.0000 52 31.9444 68.0556
28.1250 71.8750 0.0000 6776689 6776689 ACAGTCGGAGCCGCGACCGTAAGGGAGCGTCCTGCC
GGCGGGACGCTCCCTAACGGTCGCGGCTCTGAATGT
3438780 3438835 56 3438862 3438914 53 26 78.5714 5.3571 46 39.4495 60.5505
36.3636 63.6364 0.0000 6877697 6877697
GGCCCCCTTTCCCTCATCTCAGCTCTTTCTTCAGAACGGTGGGGATCGGTCAAGGG
CCCTTGACCGATTCCCTCCGTTCTGGAATCTGGTACCCGAGGGAAAGGGGGCC
3463336 3463377 42 3463426 3463467 42 48 97.6190 0.0000 79 53.5714 46.4286
53.6585 46.3415 0.0000 6926803 6926803
TTTTCTTAGCGTTACTTTGCGGCTTTGCGGCTTTGCGAGAAA

TTTCTCGCAAAGCCGCAAAGCCGCAAAGTAACGCCAAGAAAA
3548858 3548906 49 3548936 3548985 50 29 86.0000 2.0000 63 49.4949 50.5051
51.1628 48.8372 0.0000 7097842 7097843
TTTTCTCGCAGAGACGAGAGCCGAGAGAACCGCGAAGAAAAACAAAA
TTCTTCTTCTTCTTGGCGTTACTCTGCGGCTCTGCGTCTCTGCGAGAGAA
3590047 3590074 28 3590589 3590616 28 514 96.4286 0.0000 51 28.5714 71.4286
29.6296 70.3704 0.0000 7180663 7180663 CGATCGCTCCCTGACGGTCGCGGCTCTG
CAGAGCCGCCACCGTCAGGGAGCGATCG
3642325 3642379 55 3642407 3642458 52 27 80.0000 5.4545 49 42.0561 57.9439
40.9091 59.0909 0.0000 7284786 7284786
GCCCCCTTTCCTCATCTCAGCTCTTTCTTCAGAACGGTGGGAATCTGTCAAGGG
CCCTTGACGGATTCCACCGTTCTGGAATCTGGTACCCGAGGGAAAGGGGGC
3683367 3684608 1242 3687338 3688579 1242 2729 78.4455 0.9615 1127 40.1771
59.8229 39.7344 60.2656 0.0000 7371946 7371946
GGCTGACCTCCAGATAGAGCCGGAGCCGTCATCGGAAACGAGTAGCGATCCATCGGACGCAGCGGCAACTCCGA
CGGGACGGCCCCAGACGGCGCCGTTTCCAGAACGAAACCCGTGAGGAAGTCTTCATAGACTCCATTGGCTTTGC
CATCCTTGAGCGGCACGCGTATGACTTCGTAGCCGCTACGAACCGAGCGGTTCCACGAACCGTGCTGGGCCGCGA
AGATGTCCCCCTTGATTTCGGCGGGGAACTGCGCCCCCTCGTAGAACATCATCTGTAAGGACGCGTTGTGCGGTT
GCAGGAGCACATCAGGGACAATCACCTTACTCTTCAGTTCGGGATGTTTGCCGGCTTGCGGAGGATCCTGATTCC
CGCCCGTGTAAGAACCGGCCAGCCATAGAAACCGCCTCGCTCGATGTGCGTGATGTAATCCGGGACCAGATTGT
CTCCCAACCCGTCGCGTTTCATTACGGAGCACCAGAGCTCGCCGGTCGTGGGATTGACAGCCAGACCAACCGGAT
TGCGAATGCCCCGACCCGTAGACCCCGACGAACTTGCCCTCGGGTGTTGATTGAGAACATTCGCCCGATTCCCTAT
CAGCCGGATGGTCATCGGGATCGTCAGCGTTCGACTGCGACCCGAGGGCTACGAAGAGACTCTGTCCGTCGAGCG
AGAAGGCCAGATCGCGGGTGATGTGTCCTCCGGCTGGAAGGCCGGTGATGATTGTCTCTGCGTTGCCGTGGCGC
TGAGATCATCGCGCTTGATAGGCAAAATCGAAGCACGGTAGTCGTGTTGGAGACGTAGAGCCATGCCGGATTGTCGC
CTGAAGGAAAGAAAGCCATCCAAAGACCGAGCCGAGGCCGGTGCGAACACCCCTTCGAGAAGCGGTTTGCCAT
CGGCACCGGCTCCCTGTATACGCGAATCTGGCCGGAACCGGTTTCGGCCAGGAACACGTCGCCATTCCGGAGCCC
GCTTCATAGCGCGCGGCTCCGACGGCCAGAGGCGTAGAGGTTACATTGAAGCCAGGAAGGGCCTGCGGCATCA
CCCCGGCGGGCCGCGGACAACCTGCGGGAAGGCGGTAACGGACTGGGTGCGATATGGCTGTGGGAAATCCGCGG
GCGTCAACCTGCAGAAAAATCCCTGGCTTTTGTGTCTGCACGCTGCTGAATGCGGCGCTTCCCGTCAAAAATTGG
TCGGATTGGCGTCTGCCCAGACGAGCAGGCCACTTATGGCCA
TGGCCATAAGCGGCCTGCTGGCGTGGGCAGACGCTATCCGGATAAGGTGCTGACGGGAAAAGCCGCGTTACGCG
ACGTGATTACGAGAAGCCGGGCTCTTCCGGAAGCTGACGCTGCGGATCTCCAGCACCCCTATGCGACTTCGT
CCGTTAGCGGCTTTCCGCGTGTGGTCGCGCGGCCGGCTGGGGCGCTGCCGCAAGCACTGCCTGGGTTTGTGTTA
ACCTCTATGCTTCCGACCTGAAGACTCCGCGAGCCTTGAAGCGGGCTCCGAATGGGGACGTCTTCTGGCCGAAA
ACGGCAGCAGCCAAATCCGTGTGTACCGAGGGATCGGCAGCGATGGAATAATCGGCCACGGAATCTGTCTTTGCCA
CAGGTGTGCGCACGGTTTTTCGGGATGGCTTTCTATCCGCCGGGCGACGATCCCAAATGGCTCTACGTCTCGAACA
TGACCACCGTCTATCGCTTTGGGTATTCGCGCAATGATCTACTGCCACGGGTAAACCGGAGACGATCATCAGCG
GGATTCCGGGCGGAGGACATGTGACCAGGGACTTGGCCTTTTCCGAGATGGCAAGAGCCTCTTCGTGGCGATCG
GGTCTCAGTCGAATGTCAACGATCCAGATGAACGCCCCGGCCGAGAAGAACCAGCCCAATGTTCTCGAATACACAC
CCGACGGCAAGTTCGTGAGGTCTATGGTGGTGGCATCCGCAACCCCGTTGGTCTCGCGTCAACCCGACAACGG
GCGAGCTCTGGTGTCCGTGAACGAACGTGACGGGTTGGGAGATAATCTGGTGCTGATTACATCACTCACATCG
AGCCGGGCGGCTTCTATGGCTGGCCCTGGTTCTACATCGGCGGCAATCAGGATCCCCGTCTCGCCGGCAAAACCC
CCGAACCTCAAAGACAAGGTAATCGTGCCTGACGTATTGCTGCAACCGCACTTCGCGTCACTGCAGATGATTTTT
ATGAAGGCACGAGTTCCCCGCTGAATACAAGGGCGATATCTTCGCCGCGCAGCACGGATCCTGGAACCGCAGCG
TTCGAACCGGGTACGAAGTGATACGCGTACAACCTCAAAAACAGCAAGGCCAATGGAACCTATGAAGACTTCCTGA

CGGGCTTCGTCCTGGAGAACGGTGAGGTATGGGGACGTCCGGTCGGCGTCGCAGCCGCTTCCGATGGATCTCTGC
TGGTTTCTGATGACGGCTCCGGTTCCATTTGGAGAGTCAGCC
3720367 3721038 672 3728391 3729063 673 7352 98.9599 0.1486 1309 45.4275
54.5725 45.3453 54.6547 0.0000 7449429 7449429
GTGTCCTTTCGGCAAGGATGCATAGTTTGGATTTATTGGGCATTTCCGGTCTGTTCGTCTAAGGTAAACAGCTT
AGGCTGACGAATGATTATGCATGCGAACCATTGGTACTTCCGTCAAAGGAAAATCGTGTGCGCTGCTGGTTCTAC
CTGGTCAGTGCCTGCGGGTGTGAGAGCGGCGGAAAAAAGGCGAGGGTGGCTACTGCCCCACCGGACACGGCCAC
CACCAGTTTTGCGGTCGCCCTACGGAGGGCGAAGGGTCGGAGCTTACCTGTTGGCGCTGCCGATGTTCCGGTGGG
CTGGGGCAGGAACAGCGAAAATGGTTGTAGAGCGACAAGTCCATGAAGTGGGTGGCTGTGGATTGAGTGTGTTT
GCTTCGGTGGCATAACCGCTGTGACCAGCGTCAGCTGTTTCTAAGGTTCCATAGCGGCAAAGTCTATCGCTATTTC
GGGTTCCCGTCGGAACATTATGATGAGTTGCTTGACGAGAGTCTAAAGGCGGGTATTTTGCCGAGAGCATCCGG
GGGAAGTTTCTCTATGAAGAGGTTGCGGAGGCTTGACCTATTTCTTCAAGAAATAGGTCAGGCCAGAATGCGGCG
AACGGAGGTGCGACCGATCTGCAAGCGACGGGCGATCTCGGCTTTGCTGACGCCGGCAAGATACAGTTTTCG
CGAAAATGTATCTTGCCGGCGTCAGCAAAGCCGAGATCGCCCGTCGCTTGACGATCGGTGCGACCTCCGTTTCGC
CGCATTCTGGCCTGACCTATTTCTTGAAGAAATAGGTCAAGCCTCGCGAACCTCTTCATAGAGAACTTCCCCCG
GATGCTCTCGGCAAATAACCCGCTTTAGACTCTGCTGCAAGCAACTCATCATAATGTTCCGACGGGAACCCGAA
ATAGCGATAGACTTTGCCGCTATGGAACCTTAGAAACAGCTGACGCTGGTCACAGCGGTATGCCACCGAAGCAAA
CACACTCGAATCCACAGCCACCCAGTTCATGGACTTGTGCTCTACAACCATTTTCGCTGTTCTGCCCCAGCCC
ACCCGAACATCGGCAGCGCCAACAGGTAAGTCCGACCCTTCGCCCTCCATAGGGCGACCGCAAACTGGTGGTG
GCCATGTCCGGTGGGGCAGTAGCCACCCTCGCCTTTTTTCCGCCGCTCTCGACACCCGACGGCATTGACCAGGTA
GAACCAGCAGCGCACACGATTTTCTTTGACGGAAGTACCAATGGTTCGATGCATAATCATTGTCAGCCTAAG
CTGTTTACCTTAGACGAACAAACGGCGAAACGCCCAAATAAATCCAACTACGCATCCTTGCCGAAAGGACAC
3908470 3908502 33 3908505 3908537 33 2 90.9091 0.0000 51 34.8485 65.1515
36.6667 63.3333 0.0000 7817007 7817007 ATTAGCAACCGCTTCTGACGGTCGCGGCTCGG
CCGAGCCGCGCGCTCAGCAAGCGGTTGCTAAT
4028643 4028696 54 4028796 4028849 54 99 83.6364 3.6364 61 28.7037 71.2963
28.2609 71.7391 0.0000 8057492 8057492
CAGCCAATCCTGGCTGCAGCCGCTTTAGGTGGCTTCCGCGCACTCGCCCGGT
ACCGGCTCGAGCGCGGACGCCGGCTGAAAGCCGGCGGCGAGCCAGGATTGGCTG
4030038 4030067 30 4030174 4030203 30 106 93.3333 0.0000 50 25.0000 75.0000
25.0000 75.0000 0.0000 8060241 8060241 GGTGAGTTTCCCGCCGAGCCGCCCATCGC
GCGACGGGCGGCTCCGGCGGGAACTCACC
4119399 4119445 47 4119460 4119507 48 14 79.5918 6.1224 42 49.4737 50.5263
48.7179 51.2821 0.0000 8238905 8238905
GGCTAAAGAATTTCCGGCCGAGATGAACGCAGATAAACGCAGATAA
TTATCCGCGTTGATCCGTGTTTCATCCGCGGCCATTTCTGCTTTAGCC
4138197 4138224 28 4138252 4138279 28 27 89.2857 0.0000 41 41.0714 58.9286
40.0000 60.0000 0.0000 8276476 8276476 CGCAGAGCCGAGAGTTTCGAGAGAAA
TTTCTCTGCGTGAATCTGCGGGTCTGCG
4143993 4144018 26 4144043 4144068 26 24 92.3077 0.0000 42 28.8462 71.1538
29.1667 70.8333 0.0000 8288061 8288061 CAACGAGCCGCGACCGTCAGGGAGCG
CGCTCCCTTACCGTCGCGGCTCGTTG
4152598 4152639 42 4152666 4152708 43 26 88.3721 2.3256 59 43.5294 56.4706
42.1053 57.8947 0.0000 8305305 8305306
TTTCTCGAGAGCCGAGAGCCGAGAGTAACGCAGAGAGAA
TTCTTCTTGGCGTGAATCTGCGGCTCTGCGTCTCTGCGAGAAA

4152660 4152706 47 4152718 4152763 46 11 87.2340 2.1277 62 44.0860 55.9140
43.9024 56.0976 0.0000 8305424 8305424
TGGTCTTTCTTCTTGGCGTGA CTCTGCGGCTCTGCGTCTCTGCGAGA
TCTCGCAAAGCCGCCAAGCCGAGAGTAACGCCAAGAAGAAAACCA
4277430 4277470 41 4277486 4277526 41 15 92.6829 0.0000 67 45.1220 54.8780
44.7368 55.2632 0.0000 8554956 8554956
ATTTCTCGCAGAGCCGAGAGCCGAGAGTCACGCAAAGAA
TTCTTGGCGTTACTCTGCGGCTCTGCGTCTCTGCGAGAAAT
4399259 4400085 827 4408021 4408841 821 7935 68.6532 3.5757 303 38.1675
61.8325 37.1528 62.8472 0.0000 8808106 8808100
CCCATGCCCCGCGGAGCCTTTCCGTCCTCGCTGAAAGCCACGCCGCGGACGCCGACATCTTCGAACGTCCCG
TTGCGAAGATTGCGATAGAGCTTGTGGGCTGCGTGCTGTTGGCAATCAAATGTCCGGCCAGCCGTGCGGATCG
TAATCCAACAGCGCGACGCCGAGCGATTTGGACGTGCTGTCGAAGATGCCGCTCTTCGCGGTGACGTCTTCGAAG
GTGCCGTTGCCGCGGTTGCGGAACAGCCAGCAGGTCTCGCCGTGGTAGGCCTCGGGAGTGCGAACGATTTGCGC
GAGCCATCGACACTGCAAAAAACGTCGTGCTCCGCGGACCATTGACGTAATTGCAAACGAAGAGGTCCAGCAGC
CCGTCGCGATCGAAATCGAACCACATGGCCGATGTGCTGAAGGCGCTGCGGCCGCCAGGCCGGCTTTCTCGGTG
ACGTCGAGGAACCGGCCGCGCCATCGTTATGGAACAAGCGGTTCTGTCCGACGGCTGTCACCAGGAGATCTGCA
AAGCCGTCGTTGTTGTAATCGCCGATCGCGACGCCATGCCGTACATTTCCACCGCGAGCCCGGCGCGCTCGGTA
ACGTCGGTGAATGTGCCGTTGCGATTATTGCGGTAGAGCCGCAAGGTGGAGCGCTCGGTGCGATGACCGGGCCAG
CCGGTTCATTGACCAGCAAAATGTGAGCCAGCCATCGGCATCGTAATCGAAGAAGGCGCAGCCGGGGCCCATG
GTTTCGGGCAGGTACTTGGCTCCGAAGGCGCCGCTGTTGTGGTGAACTGAATGCCGGCCGCCGCGGTGACGTG
GT
ACCGATGTGACCGCGCAGGCCGGCATAACGTTACGCACAATAGCGGACGCGCCGGCAGGAAGTATCTGCCGGAA
ACTCTTGGCTCCGGCTGCGCGTTCTTCGATGTTGACGGCGATGGCTGGCCCGACATCCTGATGGTGAATAGTAAG
GACTGGACGCCGCGCGGCCGCAAATCGCTCGCGGCTCTCTACCGCAACAACCACAACGGCACCTTCACCGACATC
ACTGCGGGCAGCGGCCTGGACATCGAAATGTACGGTCTGGGCGTGGCCATCGCCGATTACGACAACGATGGCCGC
GAAGACGTTTACATCACCGCGCTCGATGGCGACCACCTCTTTCACAATGAGGGTAACGGCAAATTTGCGGATGTG
ACCAAGTCCGCCGGAATTCAGAACGCCAACTTCGGCACCCAGCGCGGCGTGGCTGGACTACGACCGCGACGGCAAG
CTGGACCTATTGTCGCCAACTATGTGCAGTGGACCCCGAAGAACGATCTCTTCTGTTGCTTATGGCGCCGCG
AAATCGTACTGCACCCCGAATCCTACAAGGTACGCGCAGCCGTCTATTCCACAACCTGGGTGGAGGCCGGTTC
GAGGACGTGCGGGGAGAAGGCGGGGATCGCGGATCCGACGAGTAAGTCGCTGGGTGTAACCGTGCTCGATTACAAC
GCCGACGGATGGCCGGACCTTTCTGTTGCCAACGACACGACGCCAACAAAGCTGTATCGCAATAACAAGAACGGC
ACCTTCACCGAAGAGGGCATGACTGCCGGCGTGGCGTACGGAGAGGACGGAGTTGCGCGCGGGGCGATGGG
4441415 4441438 24 4441465 4441488 24 26 100.0000 0.0000 48 29.1667 70.8333
29.1667 70.8333 0.0000 8882903 8882903 AACGAGCCGCGACCGTCAGGGAGC
GCTCCCTGACGGTCGCGGCTCGTT
4457683 4457703 21 4457749 4457769 21 45 100.0000 0.0000 42 52.3810 47.6190
52.3810 47.6190 0.0000 8915452 8915452 TGGTGACCTAAGTTAGGTAC
GTGACCTAACTTAGGTCACCA
4511397 4511445 49 4511476 4511524 49 30 83.6735 0.0000 58 42.8571 57.1429
39.0244 60.9756 0.0000 9022921 9022921
TTCTTCTTGGCGTTACTTTGCGGCTTTGCGGTTCTGCGAGAAATGACGC
GCGCCTTCTCTCGCAGAGCCGAGAGCCGCAAAGAACCGCCAAGCAGAA
4511484 4511515 32 4511543 4511574 32 27 87.5000 0.0000 44 35.9375 64.0625
32.1429 67.8571 0.0000 9023058 9023058 TCTCGCAGAGCCGAGAGCCGCAAAGAACCGC
GCGTTACTCTGCGGCTCTGCGTCTCTGCGAGA

4511476 4511563 88 4511634 4511721 88 70 76.4045 2.2472 69 44.3182 55.6818
39.7059 60.2941 0.0000 9023197 9023197
GCGCCTTCTCTCGCAGAGCCGAGAGCCGCAAAGAACCGCCAAGCAGAATTTGAAGCTGTTTTCTCTGCGTTACT
CTGCGGCTCTGCG
CGCAGAGCCGAGAGTGACGCAGAGAAGAATTGGCCCTTCTCTTTGCGTTACTTTGTGGCTTTGCGGCTTTG
CGAGAAATGACGC
4526153 4526211 59 4526225 4526282 58 13 86.4407 1.6949 76 47.8632 52.1368
47.0588 52.9412 0.0000 9052436 9052435
TCTTCTTGGCGTTACTTTGCGGCTTTGCGGCTTTGCGAGAAATTAATGCCCCCTTTGCGG
CCCGAACAGTGCTTCTCTCTCGCAAAGACGCAAAGCCGCAAAGTAACGCCAAGAAGA
4535664 4535704 41 4535744 4535786 43 39 84.0909 9.0909 45 51.1905 48.8095
51.3514 48.6486 0.0000 9071448 9071449
TCTCGCAGAGCCGAGAGAACCACAAGGAAACCAAGTTTT
AAAAGCTTTGGTTTTCTCTGCGTTCTCTGCGGTCTCTGCGAGA
4722441 4722529 89 4722579 4722665 87 49 77.5281 2.2472 74 50.5682 49.4318
49.2754 50.7246 0.0000 9445108 9445108
TCTCGCAACGGCGCAAAGGACGCAAAGAAGAACAGCTTCCGGCTGAAATTCTTCTTGGCGTTACTTTGCGGCTTT
GTGTCTCTGCGAGA
TCTCGCAAAGCCGCAAAGCCGCAAAGTAACGCCAAGAAGACTCCAAAATTCTTAGGTTTTCTTTGCGATCTTTGC
GCCTTTGCGAGA
4722576 4722616 41 4722627 4722668 42 10 80.9524 2.3810 42 54.2169 45.7831
55.8824 44.1176 0.0000 9445243 9445244
ATTTCTCGCAAAGCCGCAAAGCCGCAAAGTAACGCCAAGAA
TTCTTAGGTTTTCTTTGCGATCTTTGCGCCTTTGCGAGACAT
4823882 4823921 40 4823951 4823990 40 29 90.0000 0.0000 60 38.7500 61.2500
38.8889 61.1111 0.0000 9647872 9647872
TTTCTCGCAGAGCCGCGGAGCCGAGAGTCACGCCAAGAA
TTCTCTGCGTGACTCTGCGGCTCTGCGCCTCTGCGAGAAA
4841162 4841181 20 4841224 4841243 20 42 100.0000 0.0000 40 30.0000 70.0000
30.0000 70.0000 0.0000 9682405 9682405 CAACGAGCCGCGACCGTGAG
CTCACGGTCGCGGCTCGTTG
4914781 4914819 39 4914837 4914874 38 17 92.3077 2.5641 61 42.8571 57.1429
41.6667 58.3333 0.0000 9829656 9829655
TCTCGCAGAGCCGAGAGCCGAGAGTTAACGCAGAGAA
TTCTTTGCGTTACTTTGCGGCTCTGCGGCTCTGCGAGA
4925205 4925244 40 4925269 4925308 40 24 92.5000 0.0000 65 43.7500 56.2500
43.2432 56.7568 0.0000 9850513 9850513
TTTCTCGCAGAGGCGAGAGCCGAGAGTACCGCAAAGAA
TTCTTCGCGTTACTCTGCGGCTCTGCGTCTCTGCGAGAAA
4929970 4930013 44 4930045 4930088 44 31 95.4545 0.0000 78 50.0000 50.0000
50.0000 50.0000 0.0000 9860058 9860058
TCCTTTTCTCGCAAAGCCGCAAAGCCGCAAAGTAACGCCAAGAA
TTCTTTGCGTTACTTTGCGGCTTTGCGGCTTTGGCGAGAAAAGGA
4925203 4925308 106 4929974 4930086 113 4665 76.5217 9.5652 73 48.4018
51.5982 51.1364 48.8636 0.0000 9855282 9855282
CTTTTCTCGCAGAGGCGAGAGCCGAGAGTACCGCAAAGAAGAAAAAGAGCGCTGCTTGACGTGTTCTTCGCG

TTACTCTGCGGCTCTGCGTCTCTGCGAGAAA
TTTCTCGCAAAGCCGCAAAGCCGCAAAGTAACGCCAAGAAGAAGAAACCAGGATTCCTTTGTGCTGGGTTTTCT
TTGCGTTACTTTGCGGCTTTGCGGCTTGCGGAGAAAAAG
4987049 4987104 56 4987131 4987183 53 26 78.5714 5.3571 46 39.4495 60.5505
36.3636 63.6364 0.0000 9974235 9974235
GGCCCCCTTTCCCTCATCTCAGCTCTTTCTTCAGAACGGTGGGGATCGGTCAAGGG
CCCTTGACCGATTCCCTCCGTTCTGGAATCTGGTACCCGAGGGAAAGGGGGCC
5010524 5010576 53 5010603 5010658 56 26 78.5714 5.3571 46 39.4495 60.5505
36.3636 63.6364 0.0000 10021179 10021179
GGCCCCCTTTCCCTCGGTACCAGATTCCAGAACGGAGGGAATCGGTCAAGGG
CCCTTGACCGATCCCCACCGTTCTGAAGAAAGAGCTGAGATGAGGGAAAGGGGGCC
5053176 5053213 38 5053237 5053274 38 23 86.8421 0.0000 51 42.1053 57.8947
39.3939 60.6061 0.0000 10106450 10106450
TTTCTCGCAAAGACGCAGAGCCGCAGAGAACCAGCAAAG
CTTCGCGTTACTCTGCGGCTCTGCGCCTCTGCGAGAAA
5059915 5059969 55 5060273 5060327 55 303 90.9091 0.0000 85 44.5455 55.4545
44.0000 56.0000 0.0000 10120242 10120242
ATTACTGTGCATTTGTCTGCTCGCAGGCCGCAAACATCCGCACCGGAATATTTCGC
GCGAATATTCCGGTGCGGCTGTTTTCGGCTTGCGAGCGGACAAATGCACGGCAAT
5071027 5071101 75 5071210 5071284 75 108 81.3333 0.0000 80 41.3333 58.6667
39.3443 60.6557 0.0000 10142311 10142311
CTTTGCGTGACTTTGCGGCTTTGCGAGAGAGTGTAATTTCTCGCAGAGCCGAGAGCCGAGAGTCACGCAGAG
CTCTGCGGTTCTCTGCGGCTCTGCGTCTCTGCGAGAAACGGAGAATTTCTCGCAGGGCCGAGAGCCACGCAAAG
5114532 5114584 53 5114611 5114666 56 26 78.5714 5.3571 46 39.4495 60.5505
36.3636 63.6364 0.0000 10229195 10229195
GGCCCCCTTTCCCTCGGTACCAGATTCCAGAACGGAGGGAATCGGTCAAGGG
CCCTTGACCGATCCCCACCGTTCTGAAGAAAGAGCTGAGATGAGGGAAAGGGGGCC
5141542 5141567 26 5141593 5141618 26 25 92.3077 0.0000 42 46.1538 53.8462
45.8333 54.1667 0.0000 10283160 10283160 TCTTGCGGTTACTTTGCGGCTTTGCG
CGCAGAGACGCAAAGTAACGCCAAGA
5151602 5151632 31 5151659 5151689 31 26 87.0968 0.0000 42 41.9355 58.0645
40.7407 59.2593 0.0000 10303291 10303291 CGCAGAGCCGAGAGTACCGCGAAGAAGAAA
TTTCTTCTCTGCGTGACTCTGCGGCTCTGCG
5162054 5162089 36 5162113 5162148 36 23 88.8889 0.0000 52 33.3333 66.6667
31.2500 68.7500 0.0000 10324202 10324202
CCTTCCGAGCCGCGACCGTTAGGGAGCGTCCGCTAA TTAGCGACCGCTCCCTAACGGTCACGGCTCGGAGGG
5269634 5269674 41 5269703 5269744 42 28 85.7143 2.3810 52 49.3976 50.6024
50.0000 50.0000 0.0000 10539377 10539378
TTCTTGCGGTTACTTTGCGGCTTTGCAAGAAAAACGACCCT
AGGGTGATGTTTCTCGCAGAGCCGCAAAGTGACGCCAAGAA
5281225 5281257 33 5281309 5281341 33 51 84.8485 0.0000 41 46.9697 53.0303
46.4286 53.5714 0.0000 10562566 10562566 CTGAAAACTCCTTTTCTCGCAGAGCCGAGAG
CTCTGCGGCTCTGCGAGAAACAGTCTCTTTAG
5355898 5355938 41 5355950 5355990 41 11 97.5610 0.0000 77 48.7805 51.2195
50.0000 50.0000 0.0000 10711888 10711888
TTGTTCTTGCGGTTGCTTTGCGGCTTTGCGTCTTTGCGAGA

TCTCGCAAAGACGCAAAGCCGCAAAGCAACGCCAAGAAGAA
5365745 5365771 27 5365789 5365815 27 17 92.5926 0.0000 44 44.4444 55.5556
44.0000 56.0000 0.0000 10731560 10731560 CGCAGAGCCGAGAGTAACGCAAAGAA
TTCTTGGCGTTACTTTGCGGCTCTGCG
5365745 5365815 71 5366911 5366981 71 1095 79.1667 2.7778 65 45.0704
54.9296 45.6140 54.3860 0.0000 10732726 10732726
CGCAGAGCCGAGAGTAACGCAAAGAAGAAAGACCCGAGAGGTTTTCTTGGCGTTACTTTGCGGCTCTGCG
CGCAGAGCCGAGAGTAACGCAAAGAAGAAAGACCCGAGAGGTTTTCTTGGCGTTACTTTGCGGCTGTGCG
5457443 5457480 38 5457515 5457552 38 34 89.4737 0.0000 56 42.1053 57.8947
38.2353 61.7647 0.0000 10914995 10914995
TTTCTCGCAAAGCCGAGAGCCGAGAGTCACGCAAAG
CTTTGCGATTCTCTGCGGCTCTGCGGCTCTGCGAGAAA
5467940 5468023 84 5468101 5468184 84 77 76.1905 0.0000 68 44.0476 55.9524
42.1875 57.8125 0.0000 10936124 10936124
TTCTTCTTGGCGCTACTTTGCGGCTTCGCGGCTTTGTGAGAAATTGACCTTCTCTCGCAGAGACGCAGAGCCGCA
GAGTAACGC
GCGCTACTTTGCGGCTTTGCGGCTTTGCGAGAAATTGCCCGTCTCTCGCACAGACGCAGAGCCGAGAGTAACGC
CAAGAACAA
5468092 5468132 41 5468144 5468184 41 11 85.3659 0.0000 52 46.3415 53.6585
45.7143 54.2857 0.0000 10936276 10936276
TTTTTCTTGGCGCTACTTTGCGGCTTTGCGGCTTTGCGAGA
TCTCGCACAGACGCAGAGCCGAGAGTAACGCCAAGAACAA
5467940 5468184 245 5468253 5468506 254 68 69.1406 5.0781 91 44.2886
55.7114 42.9379 57.0621 0.0000 10936437 10936446
TTCTTCTTGGCGCTACTTTGCGGCTTCGCGGCTTTGTGAGAAATTGACCTTCTCTCGCAGAGACGCAGAGCCGCA
GAGTAACGCGGAGGACAACACCCAGGCCGCCAATTCTCCGGCTTCGTCAACGCAAGGGTGGTGTCTAGTGGCT
CATTTTTCTTGGCGCTACTTTGCGGCTTTGCGGCTTTGCGAGAAATTGCCCGTCTCTCGCACAGACGCAGAGCCG
CAGAGTAACGCCAAGAACAA
TTCTTCTCTGCGTGACTCTGCGGCTTTGCGGCTGTGCGAGAAATTGACCTTCTCTCGCAGAGACGCAGAGCCGCA
GAGTAACGCGGAGGACAACACCCAGGCCGCCAATTCTCAGCTTCTCTCAACGTAAGGATGGTCTCTAGTGCAG
ACCTTGTTCTATTCTTCTTTGCGTTACTCTGCGGCTCTGCGTCTCTGCGAGAAATTGCCCTTCTCTCGCACAGAC
GCAGAGCCGAGAGTAACGCGGAGAACAA
5468258 5468293 36 5468305 5468340 36 11 86.1111 0.0000 47 36.1111 63.8889
35.4839 64.5161 0.0000 10936598 10936598
CTCTGCGTGACTCTGCGGCTTTGCGGCTGTGCGAGA TCTCGCAGAGACGCAGAGCCGAGAGTAACGCGGAG
5467940 5468184 245 5468414 5468667 254 229 73.4375 5.0781 146 45.0902
54.9098 45.2128 54.7872 0.0000 10936598 10936607
TTCTTCTTGGCGCTACTTTGCGGCTTCGCGGCTTTGTGAGAAATTGACCTTCTCTCGCAGAGACGCAGAGCCGCA
GAGTAACGCGGAGGACAACACCCAGGCCGCCAATTCTCCGGCTTCGTCAACGCAAGGGTGGTGTCTAGTGGCT
CATTTTTCTTGGCGCTACTTTGCGGCTTTGCGGCTTTGCGAGAAATTGCCCGTCTCTCGCACAGACGCAGAGCCG
CAGAGTAACGCCAAGAACAA
TTCTTCTTTGCGTTACTCTGCGGCTCTGCGTCTCTGCGAGAAATTGCCCTTCTCTCGCACAGACGCAGAGCCGCA
GAGTAACGCGGAGAACAACACCCAGGCCGCCAATTCTCAGCTTCTCGTCAACCTAAGGACGATCTCTAGTGCAG
ACCTTGCTCTATTCTTCTCTGCGTTACTCTGCGGCTCTGCGTCTCTGCGAGAAATTGCCCTTCTCTCCCAAAGCC
GCAAAGCCGCAAAGTAACGCCAAGAACAA
5468092 5468336 245 5468423 5468667 245 86 73.7052 4.7809 148 45.1020

54.8980 43.2432 56.7568 0.0000 10936759 10936759
TTTTTCTTGGCGCTACTTTGCGGCTTTGCGGCTTTGCGAGAAATTGCCCGTCTCTCGCACAGACGCAGAGCCGCA
GAGTAACGCCAAGAACAACACCCAGGCCGCCCAATTCTCAGCTTTCTGTCATCCTAAGGATGGTCTCCTAGTGAG
ACCTTGTCTATTCTTCTCTGCGTGACTCTGCGGCTTTGCGGCTGTGCGAGAAATTGACCTTCTCTCGAGAGAC
GCAGAGCCGCAGAGTAACGC
GCGTTACTCTGCGGCTCTGCGTCTCTGCGAGAAATTGCCCTTCTCTCGCACAGACGCAGAGCCGCAGAGTAACGC
GGAGAACAAACACCCAGGCCGCCCAATTCTCAGCTTTCTGTCACCTAAGGACGATCTCCTAGTGAGACCTTGCTC
TATTCTTCTCTGCGTTACTCTGCGGCTCTGCGTCTCTGCGAGAAATTGCCCTTCTCTCCCAAAGCCGCAAAGCCG
CAAAGTAACGCCAAGAACA
5468414 5468454 41 5468466 5468506 41 11 90.2439 0.0000 62 43.9024 56.0976
45.9459 54.0541 0.0000 10936920 10936920
TTCTTCTTTGCGTTACTCTGCGGCTCTGCGTCTCTGCGAGA
TCTCGCACAGACGCAGAGCCGCAGAGTAACGCGGAGAACA
5468414 5468506 93 5468575 5468667 93 68 82.9787 2.1277 104 46.2366 53.7634
48.7179 51.2821 0.0000 10937081 10937081
TTCTTCTTTGCGTTACTCTGCGGCTCTGCGTCTCTGCGAGAAATTGCCCTTCTCTCGCACAGACGCAGAGCCGCA
GAGTAACGCGGAGAACA
TTCTTCTCTGCGTTACTCTGCGGCTCTGCGTCTCTGCGAGAAATTGCCCTTCTCTCCCAAAGCCGCAAAGCCGCA
AAGTAACGCCAAGAACA
5468575 5468615 41 5468627 5468667 41 11 80.4878 0.0000 42 46.3415 53.6585
48.4848 51.5152 0.0000 10937242 10937242
TTCTTCTCTGCGTTACTCTGCGGCTCTGCGTCTCTGCGAGA
TCTCCCAAAGCCGCAAAGCCGCAAAGTAACGCCAAGAACA
5554704 5554746 43 5554814 5554856 43 67 90.6977 0.0000 66 44.1860 55.8140
43.5897 56.4103 0.0000 11109560 11109560
ATTTCTCGCAGAGCCGCAGAGCCGCGGAGTCACGCAAAGAAGA
TCTTCTTGGCGTGACTTTGCGGCTTTGCGGCTCTGCGAGAAAT
5554699 5554746 48 5554948 5554994 47 201 91.6667 2.0833 74 44.2105 55.7895
43.1818 56.8182 0.0000 11109694 11109694
AAGACATTTCTCGCAGAGCCGCAGAGCCGCGGAGTCACGCAAAGAAGA
TCTTCTTTCGCGTTACTCTGCGGCTCTGCGGCTCTGCGAGAAATGCTT
5655542 5655571 30 5655678 5655707 30 106 93.3333 0.0000 50 25.0000 75.0000
25.0000 75.0000 0.0000 11311249 11311249 GGTGAGTTTCCCGCCGAGCCGCCGTCGC
GCGATGGGCGGCTGCGGCGGAACTCACC
5810658 5810685 28 5810690 5810717 28 4 89.2857 0.0000 41 35.7143 64.2857
36.0000 64.0000 0.0000 11621375 11621375 AACGAGCCGCGACCGTTGGGGAGCGCTA
TAGCGCTGCCTAACGGTCGCGGTTTCGTT
5935216 5935245 30 5935359 5935388 30 113 86.6667 0.0000 40 55.0000 45.0000
50.0000 50.0000 0.0000 11870604 11870604 GCGCAAGCTAAAGACGTTGCTGAAAAAGTC
GATTTTTTTCAGCAACGTCTAAAGCATGCGC
5951193 5951319 127 5951876 5952002 127 556 68.7500 1.5625 52 33.8583
66.1417 35.2273 64.7727 0.0000 11903195 11903195
TGCTGCTGGCGGCTTTGCCGCGATCGCCGTGTTTCTGGCCCTGGTAGGGTTGTACGGGCTACTCTTTACGCGG
TACGGCAGCGGACGGCGGAGATTGGTGTGCGGATGGCGTTGGGCGCGCAACG
CGTGCGCCCCCAACGCCGTGCGGATTCCGATCTCGTGGGTTCCGCGGCTGACGGAATACGAGAGAACCGCGTAA
ATCCCTACCGCCGCCATCAACAGCGCCAGCGCGCTCAAAGCGCCAGCAGCA

5988033 5988055 23 5988099 5988121 23 43 100.0000 0.0000 46 39.1304 60.8696
39.1304 60.8696 0.0000 11976154 11976154 TCCGTGACCAGACTCTGGTCACG
CGTGACCAGAGTCTGGTCACGGA
5988282 5988311 30 5988341 5988369 29 29 90.0000 3.3333 43 52.5424 47.4576
51.8519 48.1481 0.0000 11976652 11976652 ATCTGCGTTTTATTTGTGTTTCATCTGCGGC
GCCGCAGATGAACACAGATGAACGCAGAT
5997756 5997795 40 5997807 5997846 40 11 80.0000 0.0000 40 43.7500 56.2500
40.6250 59.3750 0.0000 11995602 11995602
TCTCGCAGAGACGCAGAGCCGCAGAGTAACGCAAAGAAGA
TCTTCTTCGCGGTTCTTTGCGGCTTTCGCGCTTTCGCGAGA
5997756 5997846 91 5997882 5997972 91 35 83.5165 0.0000 107 45.6044 54.3956
40.7895 59.2105 0.0000 11995728 11995728
TCTCGCAGAGACGCAGAGCCGCAGAGTAACGCAAAGAAGACAAGGACAAATTCTTCTTCGCGGTTCTTTGCGGCG
TTGCGGCTTTCGCGAGA
TCTCGCAAAGCCGCAAAGCCGCAAAGTAACGCAAAGAAGACAAGGGCCAATTCTTCTCTGCGGTTCTCTGCGGCT
CTGCGTCTCTGCGAGA
5997882 5997921 40 5997933 5997972 40 11 82.5000 0.0000 45 45.0000 55.0000
42.4242 57.5758 0.0000 11995854 11995854
TCTCGCAAAGCCGCAAAGCCGCAAAGTAACGCAAAGAAGA
TCTTCTCTGCGGTTCTCTGCGGCTCTGCGTCTCTGCGAGA
6474723 6474771 49 6474805 6474853 49 33 90.0000 4.0000 71 58.1633 41.8367
60.0000 40.0000 0.0000 12949576 12949576
GTTTTCTTTGCGTCCTTTGCGCCTTTGCGAGAAATATGCTTTTGATTTT
AAAGTCAGAAAGGCATTTCTCGCAAAGGCGCAAAGGACGCAAAGAAAAC
6486003 6486036 34 6486059 6486092 34 22 94.1176 0.0000 58 44.1176 55.8824
43.7500 56.2500 0.0000 12972095 12972095 TCTCGCAGAGGACGCAGAGATCGCAAAGAAAACC
GGTTTTCTCTGCGATCTCCGCGTCTCTGCGAGA
6484129 6485450 1322 6487355 6488692 1338 1904 69.8893 3.6900 570 42.5564
57.4436 42.2386 57.7614 0.0000 12972805 12972821
GGAGGAGACCCGTGTTGCCCAACTGGAACGCAAGGTCGGCCAGCAGACCGTGGAGATCGATTTTTTTGAAGGGGTG
CTTGACGCGCATCGAGGAACAGCGGATGCTGCAGGCACTGACTGGAAGCCGCGACCTGCCAGTACATCGAAGAC
CAAATCCAGAAAGGCGTCCAGATGACGATTCTGCGCATGTGTGAAGTGGGCGACATCAGCCGGGCTGGCTTCTAC
CGGTTTCGATCCGGAGGCGGAGCAGCGCGAGGAAGAGTTGGACCTGCGCGACGCGATCCAGAGGATCGCTCTGGAA
TTCCCATGCTACGGTAGGCCCAGAATCACCGCCGAGCTTCGGCGTCAGGGCTGGGCAGTGAATCATAAGAGGGTT
GGACGCATCATGCGGGAAGACAACCTCCTTTGCATTGCGCGGAGGAAGTTTGTGCTGCACCACCGATTCCAAT
CACGGCCTTCGCGTCTATCCCAACTTGGCCAAGAAGATGGAGTTGACCGGGTTGGACCAGTTGTGGGTGGCCGAC
ATCACCTACATTGGGCTGGAGACGGAGTTCATCTATCTGGCGGTGGTCTGGATGCATTCTCAAGACGGGTAATT
GGCTGGGCTCTGGGCCGACTCTGGAAGCTAGCCTGGCGGTGCAGGCGTTGAAGATGGCTTTCGGCGGCGGGGC
TCCGTCAAAGGATTGACCCATCATTCGACCGCGGCGTGCAATACGCTTCCAAGGAATACACGGATTTGCTGAAG
GAACACGGCGTTTCAGATCAGCATGAGCCGGAAGGGCAATCCCTACGACAATGCCAAATGCGAATCCTTCATGAAG
ACCTTGAAATACGAGGAGGTCTATAGACAAGAATATCGCGACCTCGCCGAGGCGCAAGCGTCGGTGGAAAAGTTC
ATTGAGAAAATCTACAACGGCAAGCGTCTGCACTCGGCGCTGGGCTACTGCTCGCCGGTCGAGTTTGAACGTTTCG
CTGGTGCCGCTGGAGCAAACCGCATGAAAAGTGAGGGCGCCCTCAGAATGCGTTTTGGGAGGCATGACGGAATCT
ATCGGTCCGATGTGTCGTTCTGCTTTTTACCCCTGGGGCGTGTCCCGCTTTCGGTCCGGCCGCTGCCAGGGTA
TAGGGCATGCCGGAAGAACATGCTCTGTCCATCGTCCGCGATGAGTTCCGGCCGGCTATTCCTTGACGGGTTGC
TCGCCAGCATTGCCCGCCTCCGCTTCACCGACACGATCAGAATAAACTATTGCCGCATCGAATGGAATGATTTA

CCATCGAACGGCGAACTGTCTCTTAACTGTTTGTCTCACCCAAGGGG
CCCCAAGGCTGAGACAACCAGAGAACCAATGATTCGCAGTTCGATGATATTTCTGTTTTTTCTGTTGGTCTAGAG
GTTTATGATGCGTGTGTGCGTGAAGCGGCGACGGGCACTGCTGGCGAGCAACCCGTTCAAGGAATAACCGATCGG
AACTCATCGCGGACGATGGGCAGGGCGTGCTCCTCCCGACGCGTCTATAGCCCGGTAACGGCGCCGACCAACGG
GAGGCGGCAGCTGGACCGGTTTATCAGATGGAATGACATATCGGACCGATAGATTCCGTCATGCCTGAGAAAAC
GCTTCCTGAGGGCGCCCTCACTTTTCATACCGCCACCTGCTGCTGAGAAGCGGTGGCCGCTAGAGAGTGTTCGAA
CTCGACCGGCGAGCGATAGCTCAAGGACGAGTGCAATCGCTTTTGGTTATAGATCTTCTCCAGGAAGGTCTTTAT
CGACGCACGGGCGTGTGCCAGATTCCGATATTCGGTGCGGTGTACCTCCTCGTATTTACGCGTTTTCATGAAGGA
CTCGCAGGCCGCAATTGTCCCAGGGTTGGCCTTGGCACTCATGCTGATTTTCGATCCCGTTGCTCTGGAGCAGTTG
CGTGAATCGTGACTGGCATACTGGACTCCGCGGTGGAGTGGTGAACCAGGCCCGGACTTACCGATCGTTGCC
AAGCGCCATGCGCAAGGCATTACGCGTGAGGGAACCTTCTAAGGTTTCCCCCAAATGCCAGCCAATGACCCGGCG
TGAGAAGGCATCCAGAATGACCGCCAGGTAGACAACTCCTCTTCCAGTCGGATGTAGGTGATATCGGCCACCCA
GAGCTGATCGACTCCGGTTAGCGTCATCTCCGCGGCCGATTCCGATATACCTTCAGACCGTGACGGGAATCGGT
GGTGGCTACGACGAACTTGCGTTTGATCACACACAGCAGGTTGTCTTCCCGCATCAGCCGCTGAATCCGCTTGCG
GTTGACCTCCCATCCACGAGCCTTGAGTTCCTCGACGATACGCCGGCTCCCATAGCAGGGCCAATCGAGGGCTAC
CCGGTGATCTCGTCGCGAAGATCCATATCGGCGGGTGCCGGCTTTTCCGGATCCAAAAGCGGTAGTATCCTGC
CCGGCTGATGCCAGCGACTCGCACATCTGCACGACACTCATGCGGGTTTCCGCCTGGACTTCTTTGTGATCTG
CTCGTAGAGGGCGCGCGTTGTTAGTGCCCCGAGCAGGCGCTGTTCTTCGACATGCTGCAAGGCCCTTTAAA
AAATCGATCTCAGGGCTTGCGGGCAACCTTACGCTCCAGTTCGGCGACCCGCGTTTCTCC
6503431 6503459 29 6503550 6503578 29 90 89.6552 0.0000 43 46.5517 53.4483
46.1538 53.8462 0.0000 13007009 13007009 CGCAGAGACGCAGAGTAACGCAAAGAAAA
TTTTCTCTGCGTGACTCTGCGGCTCTGCC
6541232 6541280 49 6541957 6542004 48 676 84.0000 6.0000 54 59.7938 40.2062
59.5238 40.4762 0.0000 13083237 13083236
TGTC AACGCAAAGTAGAAATGTCCTATTCTCTGCAAAGTAGAAATGTCC
GGACATTTCTACTTTGCGCAAATAGGACATTATCATTTTTCGCGCGACA
6615802 6615853 52 6615881 6615935 55 27 80.0000 5.4545 49 42.0561 57.9439
40.9091 59.0909 0.0000 13231734 13231734
GCCCCCTTTCCCTCGGGTACCAGATTCCAGAACGGTGGGAATCCGTCAAGGG
CCCTTGACAGATTCCACCGTTCTGAAGAAAAGCTGAGATGAGGGAAAGGGGGC
6639906 6639940 35 6639969 6640003 35 28 82.8571 0.0000 40 45.7143 54.2857
44.8276 55.1724 0.0000 13279909 13279909
TTTTCATCCGATTGATCCGTGTTTCATCCGTGGCC GGCCGCGGGTAAACACGGATCAACGCGGACAAAAA
6651158 6651270 113 6651991 6652103 113 720 98.2301 0.0000 216 38.4956
61.5044 37.8378 62.1622 0.0000 13303261 13303261
TTGTGGGGCAGCCATTCCTGGCTGTCGCGATGGCGCTGCGCGCCATTAAAGGGGATGAAAACCCCTTTGTGGGGC
AGCCAATCGTGCTGCAGCCGCTTTCAGTCGGCTTCT
AGAAGCCGCTGAAAGGCGGCTGCAGCCACGATTGGCTGCCCCACAAAGGGGTTTTTCATCCCCCTTAAATGGCGCG
CAGCGCCATCGCGACAGCCAGGATTGGCTGCCCCACAA
6667337 6667392 56 6667419 6667471 53 26 78.5714 5.3571 46 39.4495 60.5505
36.3636 63.6364 0.0000 13334811 13334811
GGCCCCCTTTCCCTCATCTCAGCTCTTTCTTCAGAACGGTGGGGATCGGTCAAGGG
CCCTTGACCGATTCCCTCCGTTCTGGAATCTGGTACCCGAGGGAAAGGGGGC
6763688 6763743 56 6763770 6763822 53 26 78.5714 5.3571 46 39.4495 60.5505
36.3636 63.6364 0.0000 13527513 13527513
GGCCCCCTTTCCCTCATCTCAGCTCTTTCTTCAGAACGGTGGGGATCGGTCAAGGG

CCCTTGACCGATTCCCTCCGTTCTGGAATCTGGTACCCGAGGGAAAGGGGGCC
6777401 6777434 34 6777457 6777490 34 22 88.2353 0.0000 48 41.1765 58.8235
40.0000 60.0000 0.0000 13554891 13554891 TCTCGCAGAGGACGCGGAGATCGCAGAGAAAACC
GGTCTTCTTTGCGATCTCTGCGTCTCCGCGAGA
6802339 6802383 45 6802410 6802455 46 26 82.9787 6.3830 48 39.5604 60.4396
41.0256 58.9744 0.0000 13604793 13604793
GATGATGCCTATACGAGCCGGGACGGTCAGGGAGCGGTCGGTAAT
ATTACCGAACGCTCCCTCACGGTCGCGGCTCGGATGAAGGCACATC
6837456 6837493 38 6837498 6837534 37 4 89.4737 2.6316 54 57.3333 42.6667
58.8235 41.1765 0.0000 13674991 13674991
GCCGGCAGATGAACACAGATAAACACAGACAAGAAAAA
TTTGTTTTGTCTGCGTTTATCTGTGTTTCTGCGGC
7046316 7046353 38 7046372 7046409 38 18 86.8421 0.0000 51 42.1053 57.8947
39.3939 60.6061 0.0000 14092725 14092725
TCTCGCAGAGCCGAGAGCCGCAAAGTAACGCAAAGAA
TTCTTCGCGGTTCTCTGCGGCTCTGCGTCTCTGCGAGA
7070773 7070792 20 7070834 7070853 20 41 100.0000 0.0000 40 30.0000 70.0000
30.0000 70.0000 0.0000 14141626 14141626 CAACGAGCCGCGACCGTGAG
CTCACGGTCGCGGCTCGTTG
7103392 7103428 37 7103451 7103487 37 22 89.1892 0.0000 54 41.8919 58.1081
42.4242 57.5758 0.0000 14206879 14206879
GTATCTCGCAGAGGACGCGAGATCGCAAACAAGACC GGTCTTCTCTGCGATCTCCGCGTCTCTGCGAGACAC
7159292 7159325 34 7159348 7159381 34 22 91.1765 0.0000 53 42.6471 57.3529
41.9355 58.0645 0.0000 14318673 14318673 TCTCGCAGAGGGCGCAGAGATCGCAAAGAAAACC
GGTTTTCTCTGCGATCTCCGCGTCTCTGCGAGA
7159292 7159381 90 7159604 7159693 90 222 78.4946 6.4516 74 42.7778 57.2222
41.0959 58.9041 0.0000 14318985 14318985
TCTCGCAGAGGGCGCAGAGATCGCAAAGAAAACCGAGGAGGCCAACCTTCTTGCTGGTTTTCTCTGCGATCTCC
GCGTCTCTGCGAGA
TCTCGCAGAGGACGCGAGAGATCGCAAAGAAAGCCGAGGAGACCCAACCTTCTTGCTGGTTTTCTCTGCGATCTCC
GCGTCTCTGCGAGA
7159604 7159637 34 7159660 7159693 34 22 91.1765 0.0000 53 42.6471 57.3529
41.9355 58.0645 0.0000 14319297 14319297 TCTCGCAGAGGACGCGAGATCGCAAAGAAAGCC
GGTTTTCTCTGCGATCTCCGCGTCTCTGCGAGA
7206128 7206170 43 7206175 7206218 44 4 84.0909 2.2727 51 48.2759 51.7241
48.6486 51.3514 0.0000 14412345 14412346
TTTCTCGCAGAGGCCGCGAGAGATCGCAGAGAAACCAAGGAGAC
GTCTTGTTTGTCTTCTCTGCGTTCTTTGCGGTCTCTGCGAGAAA
7224115 7224153 39 7224175 7224213 39 21 94.8718 0.0000 68 46.1538 53.8462
45.9459 54.0541 0.0000 14448328 14448328
TTCTCGCAGAGACGCGAGAGCCGCGAGAGTAACGCAAAGAA
TTCTTTGCGTTACTTTGCGGCTCTGCGGCTCTGCGAGAA
7248896 7248948 53 7248975 7249030 56 26 78.5714 5.3571 46 39.4495 60.5505
36.3636 63.6364 0.0000 14497923 14497923
GGCCCCCTTTCCCTCGGGTACCAGATTCCAGAACGGAGGGAATCGGTCAAGGG
CCCTTGACCGATCCCCACCGTTCTGAAGAAAGAGCTGAGATGAGGGAAAGGGGGCC

7302259 7302314 56 7302341 7302393 53 26 78.5714 5.3571 46 39.4495 60.5505
36.3636 63.6364 0.0000 14604655 14604655
GGCCCCCTTTCCCTCATCTCAGCTCTTTCTTCAGAACGGTGGGGATCGGTCAAGGG
CCCTTGACCGATTCCCTCCGTTCTGGAATCTGGTACCCGAGGGAAAGGGGGCC
7421737 7421770 34 7421793 7421826 34 22 88.2353 0.0000 48 44.1176 55.8824
43.3333 56.6667 0.0000 14843563 14843563 TCTCGCAGGGGACGCGGAGATCGCAGAGAAAACC
GGTTTTCTTTGCGATCTCTGCGTCCTTTGCGAGA
7433082 7433127 46 7433154 7433199 46 26 80.4348 0.0000 47 39.1304 60.8696
37.8378 62.1622 0.0000 14866281 14866281
GATACGCTTCGATCCGAGCCGTGACCGTTAGGGAGCGGGGGTAAT
ATTAGCGACCGCTCCCTGACGGTCACGGCTCGGTTGGCTGCTTATC
7458179 7458211 33 7458241 7458273 33 29 90.9091 0.0000 51 51.5152 48.4848
53.3333 46.6667 0.0000 14916452 14916452 ATTTCTCGCAAAGGCGCAAAGGCCGCAAAGAAA
TTTCTTTGCGCTCTTTGCGCCTTTGCGAGACAT
7464679 7464738 60 7464747 7464805 59 8 78.3333 1.6667 53 50.4202 49.5798
51.0638 48.9362 0.0000 14929485 14929485
GATTCTGGAAGGAGCAGTTTCTCGCAGAGGACGCGAGATCGCGAAGAAAAACGACAAGA
TCTTGCTGGTTTTCTCTGCGATCTCCGCGTCCTCTGCGAGACATAATTCTTCCAGACTC
7493978 7494020 43 7494037 7494080 44 16 84.0909 2.2727 51 45.9770 54.0230
45.9459 54.0541 0.0000 14988057 14988058
AAAAGCATTTCTCGCACAGACGCGAGCCGAGAGTAACGCCA
TGGAGTCTACTTTGCGGCTTTGCGTCTGTGCGAGAACCGCGTTT
7572904 7572984 81 7573111 7573191 81 126 71.9512 2.4390 45 39.5062 60.4938
38.9831 61.0169 0.0000 15146095 15146095
TCCACGTACTGCGCCACCTCCTTTGCCCGATTACGCAACGCTGACACCAAACCAGGCGCTCTTCTAGTGTGCGG
AGTTTCG
CGAACTCGCCATGATCGAGGAGCACCTGCTGTGGTGTCCACGGTGCATCGACGCAGCCGAAGAAGCGGCTCAGTA
CGTGGA
7782444 7782478 35 7782518 7782552 35 39 88.5714 0.0000 50 30.0000 70.0000
29.0323 70.9677 0.0000 15564996 15564996
CAGCCAATCCTGGCTGCCGCCGGCTTTAGCCGGC GCCGGCTGAAAGCCGGCTGTGGGCAGGATTGGCTG
7782380 7782478 99 7782609 7782707 99 130 88.1188 3.9604 134 28.7879
71.2121 26.9663 73.0337 0.0000 15565087 15565087
CGCCCCATCAGAGCCGCGACCGCAAGGGAGCGATCCCGGTTTTCGGCTCGCCCGCACTGTGGGGCAGCCAATCCT
GGCTGCCGCCGGCTTTTCAGCCGGC
GCCGGCTGAAAGCCGGCTGCAGGCAGGATTGCCTGCCCCACAGAGTGGCAGGCCGAAACTGCGGGACGCTCCCT
TACGGTCGCGGCTCTGATGGGGCG
7814231 7814266 36 7814277 7814312 36 10 97.2222 0.0000 67 51.3889 48.6111
51.4286 48.5714 0.0000 15628543 15628543
TTCTCGCAAAGGCGCAAAGGCCGCAAAGAAACAAA TTTGTTTTCTTTGCGGCCTTTGCGCCTTTGCGAGAA
7850718 7850750 33 7850773 7850805 33 22 87.8788 0.0000 46 42.4242 57.5758
41.3793 58.6207 0.0000 15701523 15701523 CTCGAGAGAACGCAGAGCTCGCAAAGAAAACC
GGTTTTCTCTGCGATCTCCGCGTCCTCTGCGAG
7926155 7926194 40 7926212 7926253 42 17 90.4762 4.7619 60 43.9024 56.0976
42.1053 57.8947 0.0000 15852406 15852408
TTTCTCGCAGAGCCGAGAGCCGAGAGTAACGCCAAGAA

TTCTTGCCTTGTCTCTGCGGCTCTGCGGCTCTGCGAGAAA
7934575 7934616 42 7934629 7934669 41 12 83.3333 2.3810 47 43.3735 56.6265
42.8571 57.1429 0.0000 15869245 15869244
GGTTTTCTTCGCGTACTCTGCGTCTCTGCGTCTCTGCGAGA
TCTCCCAGAGACGCGATCTGCAGAGTCACGCAAAGAAGACC
7974180 7974235 56 7974262 7974314 53 26 78.5714 5.3571 46 39.4495 60.5505
36.3636 63.6364 0.0000 15948497 15948497
GGCCCCCTTTCCCTCATCTCAGCTCTTTCTTCAGAACGGTGGGGATCGGTCAAGGG
CCCTTGACCGATTCCCTCCGTTCTGGAATCTGGTACCCGAGGGAAAGGGGGCC
7979772 7979815 44 7979857 7979899 43 41 86.3636 2.2727 56 43.6782 56.3218
42.1053 57.8947 0.0000 15959672 15959672
TCAGTCTCTCGCAGAGACGAGAGCCGCAAAGTAACGCAGAGAA
TTCTTTGCGGTTCTCTGCGGCTCTGCGGCTCTGCGAGAGATGA
7979857 7979897 41 7979903 7979943 41 5 80.4878 0.0000 42 43.9024 56.0976
42.4242 57.5758 0.0000 15959800 15959800
TTCTTTGCGGTTCTCTGCGGCTCTGCGGCTCTGCGAGAGAT
ATTTCTCGCAAGCCGCAAAGCCGCAAAGTAGCGCCAAGAA
8116519 8116574 56 8116601 8116653 53 26 78.5714 5.3571 46 39.4495 60.5505
36.3636 63.6364 0.0000 16233175 16233175
GGCCCCCTTTCCCTCATCTCAGCTCTTTCTTCAGAACGGTGGGGATCGGTCAAGGG
CCCTTGACCGATTCCCTCCGTTCTGGAATCTGGTACCCGAGGGAAAGGGGGCC
8159670 8159703 34 8159726 8159759 34 22 97.0588 0.0000 63 42.6471 57.3529
42.4242 57.5758 0.0000 16319429 16319429 TCTCGCAGAGGACGCGGAGATCGCAAAGAAAACC
GGTTTTCTCTGCGATCTCCGCGTCTCTGCGAGA
8320614 8320647 34 8320670 8320703 34 22 94.1176 0.0000 58 44.1176 55.8824
43.7500 56.2500 0.0000 16641317 16641317 TCTCGCAGAGGACGCGGAGATCGCAGAGAAAACC
GGTTTTCTTTGCGATCTCTGCGTCTCTGCGAGA
8333953 8333977 25 8333995 8334019 25 17 92.0000 0.0000 40 52.0000 48.0000
52.1739 47.8261 0.0000 16667972 16667972 TTATCTGCGTTTATCTGCGGCCCAT
ATGGGCCACAGATGAACGCAGATAA
8390533 8390588 56 8390615 8390667 53 26 78.5714 5.3571 46 39.4495 60.5505
36.3636 63.6364 0.0000 16781203 16781203
GGCCCCCTTTCCCTCATCTCAGCTCTTTCTTCAGAACGGTGGGGATCGGTCAAGGG
CCCTTGACCGATTCCCTCCGTTCTGGAATCTGGTACCCGAGGGAAAGGGGGCC
8395365 8395398 34 8395402 8395435 34 3 91.1765 0.0000 53 36.7647 63.2353
35.4839 64.5161 0.0000 16790800 16790800 TCCGAGCCGTGACCGTCAGGGAGCGGTTGGTAAT
ATTACCGACCGCTCCCTGACGGTCGCGGCTCTGA
8521119 8521147 29 8521184 8521212 29 36 93.1034 0.0000 48 50.0000 50.0000
51.8519 48.1481 0.0000 17042331 17042331 ACGTTAGATCTCTATCGGAAGTGACGCGG
CCGCGTCAGTTTCGATAGAGATCTAACGT
8743822 8743859 38 8743861 8743898 38 1 86.8421 0.0000 51 31.5789 68.4211
30.3030 69.6970 0.0000 17487720 17487720
CTCCATCCGAGCCGCGGCCGTAAGGGAGCGGTCGCTAA
TTACCGGACGCTCCCTGACGGTCGCGGCTCGGATGGAG
8743748 8743778 31 8743869 8743900 32 90 90.6250 3.1250 47 28.5714 71.4286
27.5862 72.4138 0.0000 17487647 17487647 GTCCCATCCGAGCCGCGACGGTTAGGGAGCG

CGCTCCCTGACGGTCGCGGCTCGGATGGAGAC
8766583 8766617 35 8766678 8766712 35 60 91.4286 0.0000 55 55.7143 44.2857
56.2500 43.7500 0.0000 17533295 17533295
GTTTTCTTTGCGTCTTTGCGCCTTTGCGAGACAT ATATCTAGCAAAGGCGCAAAGGACGCAAAGAAAAC
8779504 8779547 44 8779584 8779627 44 36 93.1818 0.0000 73 56.8182 43.1818
56.0976 43.9024 0.0000 17559131 17559131
AAGATTTTCTTGCGATCTTGCGTCTTGCGGAGAAAAGCTTTT
AAAAACATTTCTCGCAAGACGCCAAGACCGCCAAGAAAATCTT
8791320 8791355 36 8791407 8791442 36 51 88.8889 0.0000 52 48.6111 51.3889
50.0000 50.0000 0.0000 17582762 17582762
TTAGCTTTTCTTTGCGGTCTTGCGTCTTGCGGAGA TCTCGCCAAGCCGCAAAGACCGCCAAGAAAACCTAA
8818171 8818210 40 8818230 8818268 39 19 90.0000 2.5000 58 54.4304 45.5696
55.5556 44.4444 0.0000 17636440 17636440
GTTCTTTATCTGCGTTCATCTGCGTTCATCTGTGGCCCAT
ATGGGCCCGCAGATAAACGCAGATGAACACAGATAAAAAAC
8856602 8856653 52 8856681 8856735 55 27 80.0000 5.4545 49 42.0561 57.9439
40.9091 59.0909 0.0000 17713334 17713334
GCCCCCTTTCCCTCGGGTACCAGATTCCAGAACGGTGGGAATCCGTCAAGGG
CCCTTGACAGATTCCACCGTTCTGAAGAAAAGAGCTGAGATGAGGGAAAGGGGGC
8943875 8943910 36 8943922 8943957 36 11 83.3333 0.0000 42 38.8889 61.1111
36.6667 63.3333 0.0000 17887832 17887832
CTTGCGTACTCTGCGGCTCTGCGTCTTTGCGCGA TCTCGCAACGCCGCAAAGCCGCAAAGCAACGCCAAG
8989981 8990037 57 8990047 8990100 54 9 78.9474 5.2632 48 54.0541 45.9459
57.7778 42.2222 0.0000 17980084 17980084
TATTCGTTCTGGCTTTTCTTGTTGGCCTTACTTTGCGGCTTTGCGGCTTTGCGATAA
TTCTCGCAGAGTCGCAAAGCCGCAAAGTAACGATAAGAGAACCTAGGACGAATA
9009795 9009867 73 9009868 9009942 75 0 74.0260 7.7922 42 41.8919 58.1081
38.5965 61.4035 0.0000 18019735 18019738
GCTAAGCTAAGTACACCGTTTCAGAAGTTCGATCGCGTATCGCGCGATCGCTTGCTGACGCGCGCGGCTCCGA
TCAGAGCCGCGACCATTAGAGAGCGATCGCCAGCAAGTGCAGCGCTGTACTTCTGAAACGGCGTACCAAGCTAGC
9040603 9040664 62 9040780 9040841 62 115 76.1905 3.1746 47 41.1290 58.8710
37.5000 62.5000 0.0000 18081444 18081444
CTTTGCGTGACTCTGCGGCTCTGCGGCTTTGCGAGAAATCTGCCATCTCTCGCAACGCCGCA
TGCGGCTCTGCGAGAAAAGCACTGTCTCTCGCAACGCCGCAAAGCCGCAAAGTATCGCCAAG
9040759 9040794 36 9040806 9040841 36 11 86.1111 0.0000 47 40.2778 59.7222
38.7097 61.2903 0.0000 18081600 18081600
CTTTGCGATACTCTGCGGCTCTGCGGCTCTGCGAGA TCTCGCAACGCCGCAAAGCCGCAAAGTATCGCCAAG
9088174 9088219 46 9088222 9088268 47 2 80.8511 2.1277 47 58.0645 41.9355
57.8947 42.1053 0.0000 18176441 18176442
TGGCCGCAGATGAACGCAGATAAACACAGATAAGACAAGAACAAGA
TCTTGTTTTATGTGTTCTCAGCGTTTATCAGTGTTTATCTGCGGCCA
9101605 9101647 43 9101652 9101694 43 4 86.0465 0.0000 56 53.4884 46.5116
54.0541 45.9459 0.0000 18203299 18203299
TGGGCCGCAGATGAACACAGATAAACGCAGATAAAAACAAAAGC
GCCTTTGTCTTATCTGTGTTTCATCTGCGTTTATCTGCGGCCCA
9109130 9109176 47 9109178 9109223 46 1 87.2340 2.1277 62 54.8387 45.1613

53.6585 46.3415 0.0000 18218354 18218353
CTTATCCGCGTTTATCCGTGTTTATCCGCGGCCAAATTTGCTTTTC
GAAAATCAATATGGGCCACGGATGAACACGGATAAACGCGGATAAG
9175236 9175271 36 9175284 9175320 37 12 86.4865 2.7027 47 32.8767 67.1233
31.2500 68.7500 0.0000 18350555 18350555
GCCCCATCCGAGCCGCGACCGTTAGGGAGCGATCT AGATCGCTCCCTGAGGGTCGCGGCTCTGATAGAGGGC
9212119 9212144 26 9212238 9212263 26 93 92.3077 0.0000 42 26.9231 73.0769
25.0000 75.0000 0.0000 18424382 18424382 GAGCCGCGACCGCAAGGGAGCGATCG
CGATCGCTCCCTCACGGTCGCGGCTC
9244800 9244849 50 9244940 9244989 50 90 94.0000 0.0000 85 51.0000 49.0000
51.0638 48.9362 0.0000 18489789 18489789
GAAGGGCTCATTTCTCGCAAAGACGCAAAGCCGCAAAGTAACGCCAAGAA
TTCTTTGCGTTCCTTTGCGGCTTTGCGTCTTTGCGAGAAATGAGCCTTTTC
9264052 9264078 27 9264112 9264138 27 33 92.5926 0.0000 44 46.2963 53.7037
48.0000 52.0000 0.0000 18528190 18528190 TTCTTCGCGTTGCTTTGCGGCTTTGCG
CGCAAAGCCGCAAAGTAACGCCAAGAA
9284181 9284210 30 9284234 9284263 30 23 86.6667 0.0000 40 48.3333 51.6667
46.1538 53.8462 0.0000 18568444 18568444 TTCTTCTAGCGATGCTGTGCGGCTTTGCG
CGCAAAGCCGCAAAGTAACGCCAAGAAGAA
9284203 9284271 69 9284350 9284416 67 78 86.9565 2.8986 89 44.1176 55.8824
43.3333 56.6667 0.0000 18568621 18568620
GCTTTGCGAGGGCGGCTGATTCTCGCAAAGCCGCAAAGCCGCAAAGTAACGCCAAGAAGAAAGGACAAG
CTTGTTGCTCTTCTTGGCGTTACTTTGCGGCTTTGCGGCTTTGCGAGAACCCGTAGCCTCGCAAGGC
9336725 9336755 31 9336825 9336855 31 69 96.7742 0.0000 57 54.8387 45.1613
53.3333 46.6667 0.0000 18673580 18673580 TTTCTTCATTGCGTTACTTTGCGGCTTTGCG
CGCAAAGCCGCAAAGTAACGCCAAGAAGAAA
9554413 9554443 31 9554472 9554502 31 28 93.5484 0.0000 52 33.8710 66.1290
34.4828 65.5172 0.0000 19108915 19108915 CATTCCGGGGCGAAGCCGAACACCATTCCGG
CCGGAACGCTGTTGCGCTTCGCCCCGGAATG
9554338 9554394 57 9554453 9554507 55 58 77.1930 3.5088 45 41.9643 58.0357
43.1818 56.8182 0.0000 19108847 19108846
GAACACGATTCCGGGCTGAAGGCGAACACCATTCCGGGATGAAGCCGAACAGGGTTC
GAACACTTTTCGGCGAACCCCGGAACGCTGTTGCGCTTCGCCCCGGAATGTTTTTC
9554368 9554465 98 9557062 9557157 96 2596 77.5510 2.0408 82 42.2680
57.7320 40.7895 59.2105 0.0000 19111527 19111527
ATTCCGGGATGAAGCCGAACAGGGTTCCGGGATGATTCCGAACGGCATTCCGGGGCGAAGCCGAACACCATTCCG
GCCTGAAGCCGAACACTTTTCGG
CCGAATGGTGTTCGGCATCATCCCGGAATGCTGTTGCGCTTCATCCCGGAATGGTGTTCGACTTCCCTCCGGAAT
GCTGTTGACTTCGCCGGAAT
9554317 9554428 112 9557054 9557163 110 2625 83.0357 1.7857 125 40.5405
59.4595 41.9355 58.0645 0.0000 19111482 19111481
GTGCGTATTCCGGGGCAAACCGAACACGATTCCGGGCTGAAGGCGAACACCATTCCGGGATGAAGCCGAACAGGG
TTCCGGGATGATTCCGAACGGCATTCCGGGGCGAAGC
GCTACGCTCCGAATGGTGTTCGGCATCATCCCGGAATGCTGTTGCGCTTCATCCCGGAATGGTGTTCGACTTCCC
TCCGGAATGCTGTTGACTTCGCCGGAATCCGCAC
9612981 9613009 29 9613148 9613178 31 138 90.3226 6.4516 43 36.6667 63.3333

35.7143 64.2857 0.0000 19226157 19226158 TTCCCTGTCCTGCCCGTTGGTCTCGGCAC
 GTGCCCCGAGACCAACGGGCAGAACAGGAGAA
 9630087 9630123 37 9630135 9630172 38 11 86.8421 2.6316 49 46.6667 53.3333
 45.4545 54.5455 0.0000 19260258 19260258
 TTCTTGCGGTTACTTGCGGCTTTGCGTCGTTGCGAGA
 TCTCGCAGAGACGCAAAGCCGCAGAGTTACGCTAAGAA
 9630059 9630174 116 9630302 9630419 118 127 79.8319 3.3613 110 50.0000
 50.0000 50.5263 49.4737 0.0000 19260476 19260476
 GTAAGAACGCTAATTCGTTCTTGTGTGTTCTTGCGGTTACTTGCGGCTTTGCGTCGTTGCGAGAAAGCTGCAGT
 CTCTCGCAGAGACGCAAAGCCGCAGAGTTACGCTAAGAAGA
 TCTTCTTGCGGTTACTTTGCGGCTTTGCGTCTTTGCGAGAAAGCTGCAGTCTCTCGCAAAGACGCAGAGCCGCGG
 AGTTACGCCGAGAAGAAAAGCATAACGAAGAACTTTGCTTAC
 9630290 9630341 52 9630353 9630404 52 11 83.0189 3.7736 57 48.0769 51.9231
 45.4545 54.5455 0.0000 19260694 19260694
 TCGTTCTTGCTCTTCTTGCGGTTACTTTGCGGCTTTGCGTCTTTGCGAGA
 TCTCGCAAAGACGCAGAGCCGCGGAGTTACGCCGAGAAGAAAAGCATAACGA
 9742675 9742713 39 9742730 9742768 39 16 87.1795 0.0000 53 50.0000 50.0000
 50.0000 50.0000 0.0000 19485443 19485443
 TTCTCGCAGAGCCGCAAAGCCGCAAAGTAGCACAAAGAA
 TTCTTTGTAGTTCTTTGCGGCTCTGCGGCTTTGCGAGAA
 9752253 9752287 35 9752308 9752342 35 20 94.2857 0.0000 60 45.7143 54.2857
 45.4545 54.5455 0.0000 19504595 19504595
 TCTCGCAGAGGACGCAGAGATCGCAAAGAAAACCA TGGTTTTCTCTGCGATCTCCGCGTCCTCTGCGAGA
 9765730 9765766 37 9765807 9765843 37 40 91.8919 0.0000 59 25.6757 74.3243
 26.4706 73.5294 0.0000 19531573 19531573
 CCAACGAGCCGCGACCGTCAGGGAGCGCCGAGCCCTC GACGGCTCGACGCTCCCTCACGGTCGCGGCTCGTTGG
 9921348 9921374 27 9921395 9921421 27 20 100.0000 0.0000 54 44.4444 55.5556
 44.4444 55.5556 0.0000 19842769 19842769 CGCAGAGCCGCAGAGTAACGCAAAGAA
 TTCTTTGCGTTACTCTGCGGCTCTGCG
 9921343 9921362 20 9921415 9921434 20 52 100.0000 0.0000 40 40.0000 60.0000
 40.0000 60.0000 0.0000 19842777 19842777 TTTCTCGCAGAGCCGCAGAG
 CTCTGCGGCTCTGCGAGAAA
 9955126 9955161 36 9955184 9955219 36 22 91.6667 0.0000 57 54.1667 45.8333
 54.5455 45.4545 0.0000 19910345 19910345
 CTTATCTGCGTTTATCTGCGTTTATCTGCGGCAAT ATTGGCCGCAGATGAACACAGATGAACGCAGATAAG

Sequence: gil94967031|refINC_008009.1| Acidobacteria bacterium Ellin345,
 complete genome

Parameters: 2 3 5 80 10 40 500000 10000 -d -h -t4 74 -t5 493 -t7 10000

68737 68790 54 68894 68949 56 103 80.3571 3.5714 53 32.7273 67.2727 31.1111
 68.8889 0.0000 137684 137686

TGGGTGCCCCGGGTCTCGCGATTTTCGAGACTTGGGAGAGAGGATGTCTGCACC
GGTGCCGGGATTCTCGTCCGCCAGGTCTCGAACCCCGCGAGACCCGGGGCACCCA
143424 143470 47 143488 143534 47 17 80.8511 0.0000 49 58.5106 41.4894
60.5263 39.4737 0.0000 286958 286958
AAAGCCAAAATCTTTGAGCACGAAGGACACGAAGGTTTCACGAAGGA
TCCTTTGTGGGAACCTTCGTGTCCTTCGTGTTAGAGATTTTTCTTT
499604 500967 1364 506610 507989 1380 5642 67.2792 6.0777 343 43.3673
56.6327 43.1723 56.8277 0.0000 1007577 1007598
TTGCCGCTGCGGTGAAAAACGAGCCAACATAATCCGCGAGCCAGAGCGCTCCGAAGATGTCTACCGACGACTCG
CCACTGTTCCACGTGATGTTTCGATTCGGTAACGAACATTGGCACATCGGCAGGCAGACCGTCGTCACGCCAAACC
TGCATGATGTGAGTCATGAGCTGCGGCTCGTCGTAAGGCTGCTCCAATTCACCTTGACGGTTCGAATGGATAG
TGTTTGAACGAGAAGAAGGCGAGCTCTCGCAGACCGCCGCGGGCCCTCAAATAGTTTCAGGAACCGGCGCGTCCAC
GATGTCTGACCTTGCGCATCGGGCCACGACAAAATGTCTTCGTTCTGGCCCGTAAATACCGGCCCGCCAAGTTTC
AGCTTCGGATCCACCTTGTTGGATTGCCCTCGCCATTGCAGGTAGAGTTCAGCGTCGTCTTCGGGAAGCGTGTAC
TGGCCATCCGGCTCTTCGCCGAGTTCACATAGGAGATCGGGTAGCCGCGCTTCTCGACGTACGCGATTTTCAGCA
GCGGAATCTTCGGGCGTGCCATAGAGCAACGCGGTTGGAATCATTGCCGGCAAGCCGCGTGTGTATCCGCTGCGA
TAAACAGATCCATGCCAACGTGTTCTGTGTTGCGTCGATCCCAAGTCCGAAGGCTCGTGCCAAGGGTCAATCGAA
GAGCAGTAAGTCGTGGTCTGATCCTGGTCGGGGGTGTGCCGCGCCACATCATGGAAGCCATTCTGTGCGGTGCGC
CCGATGAACACCTCGCGGATTGCATAACCAACGCAATTGCGAATATCGGCCGAGCCATGTGTGTGCGAGGTATTC
GATGGCTCCGTCATCCAGACGCGCAAGAAACGCACCGGCATCGGCCAATTCGAGAGTTGGTGGGTGACCGTCCCA
CCCTTGCCGGCGCTTACCACGCCCCGCGGGGAACGCGACCCACGATCCTTTGCTCGGTAATTTGATCGCGTCGTCG
CCGGTCCAATACTGAATTAGGTATTGCTTGGCGTAAGGCTCTGCCAGGCGATCCGCAAAGCGTTTATGGATTGT
GTCGTTGCAAGATCAATCACGATCCACTGCGGATGTTCCGAATCTTCTGGCCAGTGAAATGTTTCGTGAGATAC
GGATTGCTCTTCCAATAGGTGTTGAGATCGCCGTCGGTCATTCTACCGAAGCCGCTCGTGTGCGTTCCATCGTTA
CGCGTAAAGCCACGATGGGGCAAAACATATCCGTACGAATGCTTAATGATCTCCGGTCCCGGTTCCGGCGCTACCG
GTGAAGTATCCTTTGCCGCTCGGATCGCTCCACGTGCCCTGCGGATTCCAGTGCCAGGCCTCAACTGCCAACTCG
GTGTTCTGCCGATA
TATCGCCAGAACACTGAGCTCTTCATTCAAGCGTGGCATTGGAACCCTAAGGGCAAGTGGAGCGATCCGAGCGGC
AAGGGCTATTTCTGTCGGCGACGCCACGCCTACGAGTGAGCCGATTCGCCACTCGTTCGGATACTCGCTCCAGCAT
CGTGGCTTTACCCGCAACGGGGGCTCCGAGTTCGACGGCTTCTCTCGTCTTGATGATGGCGACATCAAACTTAT
TGGAAGAGCAATCCCTACCTTACACAGACGTTCACTGGCGAACC GGATTCCGCGCATGCGCAATGGATCGTGATT
GATTTGAAAAGCCGAGGACATCAACGCGATCCGATCGCATGGGCCGAACCGTATGCGCGCGAATACAACGTG
CAGTACTGGGAAGGTTCCGGGCGACGCGATGGATGAACAGGATAAAGGCACGTGGAAGAATTTTCATCTGGGGCT
GTCGCGAACGGTGAGGGTGGCACTCCGACGATCAAGCTCAGCACCTCAACGGTCCAAGCGCGATACGTGCGTGTG
CTGATGACCGGTTCTCCAATACCTGCGACACCCACGGATCCAGTGACAAGAGAAATTGTGTTGGTTACGCCATT
CGCGAGGTCTATCTGGGTACAGTCGATCAGAGCGGCGATTTCCACGACCTGCTCAAACACACTCCGGGACAGCAG
CAAACATTGACGTATTGCTCCTCCGTTGATCCCTGGCATGAACCGTCCGACCTGTACGTGCGACCAGACCGGATG
GAATCGGGAGATCAGCCCGGCTTCGACTTGTTCTTACCAGCGGCATCACCCGCGGACTTCCTGCAATCGTACCG
ATTGCGTTGATCTATGGAACGCCAGAAGATTCCGCGGCCCAAATGGCTTACCTCAAAGCTCGCAAATATCCGATC
TCATACATCGAGATGGGTGAAGAGGCCGATGGTCAGTACATGCAGCCCGAGGACAACGCGGCTCTGTACATTAG
TGGGCGACAGCCATTACAAGGTCGACCCACATTCAAGCTGGGCGGCGGCTGTTCCAGGGCGTGACTGAAGAC
ATCAAGGCGTGGGCTGACCCGAAGGGCAGAACCTCGTGGTTCGCGCGTTTCCTTGACTACTTAAAAACCCACGGC
CATCTCGACGATTTTCGATTATGTCTGTTTCAACACTATCCCTATGACGGATGCGAAACGCCGTGGGAGAACCTT
TATCAGGAACCTGAGCTAATCATGCAGTTATGGATGTCTGGCGAGCGGATGGATTGCCGCCAAATATTCGCGTG
CTCGACACCGAAACCAACGATCACGGTGGGGAAGCGGCGGTGGACATCTTCGGCGCGCTTTGGCTCGGCGATTCT
TTCGCCGTTTCTTACC GCCGCGGACAA

865042 865073 32 865085 865117 33 11 90.9091 3.0303 49 50.7692 49.2308
50.0000 50.0000 0.0000 1730158 1730159 AACACGAAGGGCACGAAGGAAACGAAGTTTC
GAAATCCTTCGTGTACTTCGTGCCCTTCGTGT
910438 910538 101 910931 911031 101 392 72.2772 0.0000 62 41.0891 58.9109
39.7260 60.2740 0.0000 1821469 1821469
GGGCCTCTCTACAACAGTTTCGATTGGGGCGGATTCTTGATCTGGTACCTGCCGCAGTATCCGGTTTCGGTCGAT
GGCCGAAACGATCTTTACGGAGACGA
TCATCGCCGTAGAGATTGGCTCGGCCGTCGATCGACACGGGCTGCTCCGGCAAGTACCACCGGATGAACCCTCCC
CACTCCAGGCTGTTGAACAGTGGCCC
1131165 1131185 21 1131186 1131206 21 0 100.0000 0.0000 42 33.3333 66.6667
33.3333 66.6667 0.0000 2262371 2262371 GACGAGTACAAGGGGCTGCCG
CGGCAGCCCCCTTGACTCGTC
1162731 1162797 67 1163282 1163348 67 484 74.6269 0.0000 49 50.7463 49.2537
52.0000 48.0000 0.0000 2326079 2326079
AATCACTGGATCGCCTTTAAGGCAATCGGCACGAAATCAAACCGCAGCGGCATCGGAGCGAAGATTC
GAATCCTTGTCCTCAATTCCGCTGCGGTTTCGATTTCGTACCCACAGTCTTGAACCTGATCCAGTTATT
1161606 1162322 717 1163757 1164470 714 1434 67.9063 2.8926 245 40.8106
59.1894 42.3935 57.6065 0.0000 2326079 2326076
CTCTACCACAACAACCACGACGGCACCTTCACGGATGTCACCAAGAAGGCGAACCTCTGGCATGCAGGATGGGGC
CAAGGCGCCTGCGTCGGCGACTACGACAACGACGGCAACGACGACCTCTTCGTACCTACTACGGCAAGAACATC
CTCTACCACAACAACGGAAACGGAACATTACCGACGTTAGCGAGAAAGCTGGCGTCGCCGGCAGCGGCAAAATG
TGGGGAACGGGCTGCGCCTTCGTGATTACGATCGCGACGGCAAACCTCGACCTCTTCGTGCGCAATTACGTGGAC
TTCGATCTCTCTACCGCGCCAGCTCCCGCGAGCGTCCACGTGCATCTGGAAGGGTGACCGGTGATGTGCGGA
CCGCGCGGCCTGCCCTGGACTAAGAACCTGCTGTACCACAACAAGGGCGACGGCACCTTCGAAGACGTGACCGAC
AAGGCGCGCATCGATAAGACCAATGGCCATTACTCGTTCAGCGTCTCAGTCGTTGATTACGACGACGACGGCTGG
CCCGACATCTTCGTGCGCTGCGACAGCACCCCGCCATCCTGTACCACAACAACCACGACGGCACGTTTACCGAC
GTTGCTGTACCCCTCGGCGCAGCCTTCAACGAAGACGGCGCGAGCAGGCCGGCATGGGTTCCACCGTCGCCGAT
TATGACGGCGATGGCCGTCTCGACATCTTCAAAACCAACTTC
GAAGTTTGTCTTCACGATGTCGAGATTGCCATCCAGATCGTAATCCGCCGTGGAAACGCCCATGCCCGCTGCGG
TTTGCCGTCCGGGCTGAACGCGCACCCCGCATCCACCGCAACATCCTCGAATTTCCCGTTGTGCTTGTCTGATA
GAGCGCACTCGGCATCGAGTCATTGCGCCACGTAGATATCGGGCCAGCCATCGTTGTCAAATCCGCCACGCACAC
TCCGAGTCCGTAAGTCCCACACCCCTTCGTGATCCCGGCCGCTCACTCACATCGGTAAACGTCCCGTCGCCGTT
GTTGTGGTACAGGATGTCCGTATCGCCCTGCATCCCCGGGGGGCCGCATGCGACCATAATGCCCTTGTACAAGCA
CGGACCCGATTCCGGCACCGGCGCGGTTTTCAAATCCAAATCAAGATAGTTCGCGACGAAGATATCCAGATGTCC
ATCACGGTCGTAATCCACAACGCGCATCCGGTTCCCCAGCGCGTGCGCTTACCCGCGATCCCCGCCTTCTCGCT
GACATCGGCGAAGGTTCCATTGCCGTTGTTGTGATAGAGAACGTTCTTCCCGTAATACGTGACGAAAAGGTCGTC
GTTGCCGTGTTGTCGTAGTCGCCCACGCAGCAACCTGCCCCCAGCCCGATCGTCGTAGGCCTGCCTTGTGCGT
CACGTCTGTAAACGTACCGTCTCTATTGTTCTTGAAGAG
1475126 1475154 29 1475176 1475204 29 21 89.6552 0.0000 43 43.1034 56.8966
42.3077 57.6923 0.0000 2950330 2950330 ACCACGAAGGACACAAAGGACCACGAAGG
CCTTCGTGATCCTTCGTGCCCTTCGTGGT
1572718 1573822 1105 1582591 1583684 1094 8768 67.4956 4.7069 316 40.7003
59.2997 41.9737 58.0263 0.0000 3156413 3156414
CTACCAGCTTGGCATTGATGGCATCTCGTTTCTGCTGATCATCCTGACGACGTTGCTGGGTTCGATTCGATCCT
GTCGTCTGGACGGCGATTGAGAACC GCGTGAAGGAATACTACGCATGGTTCCTCCTGTTGCAGACCGGCATGCT
CGGCGTCTTCATGGCCAACGATTTCTTCTATTCTTCGTGTTCTGGGAAGCGATGCTCGTGCCCATGTACCTGCT

GATCGGCATTTGGGGCGGGCCGCGCAAGCTGTACGCCGCGATCAAGTTCTTCTCTACACGCTGGCCGGCTCGGT
 GCTGATGCTGCTGGGCATCCTGTTCTGTACTTCCACCACCACACAGTGACGGGCGTGACACGTTCTCGCTGGA
 AGCGCTCTATGCGACGGCACCGCTGATTACGGGCAAGGCCGCGATCTTCTGTTCTTCGCATTCTTCGTCGGCTT
 CGCGATCAAAGTGCCGATGTTCCCGTTCCACACCTGGCTGCCGGATGCGCACGTAGAAGCGCCAACGGCCGGCTC
 GGTGATCCTGGCGGGCGTTCTGTTGAAGATGGGAACGTACGGCTTGATCCGCTTCTCGCTGCCGTTTTTCCCGGC
 GGTGGCGACGATCAAATGGGTTCAGGCGTTGTTGATTACGCTGTCGCTGATCGGCATCCTCTACGGCGCGCTGGT
 TTCGCTGGTGCAGAAAGACATGAAGAAGCTGGTGGCGTATTCTGTCGGTCAGCCACTTGGGCTTCTGCACCCTGGG
 CATCTTCGTGTTACGCGAGATGGGTATTGCCGGTGGTGGTATCAACAGATTAACCACGGTATCTCGACTGGTGC
 ACTGTTCTTGATTGTCGGCATTGTTGTACGAGCGGGCGCATACGCGTGAGATCTCGGAATACGGTGGGATCAGTAC
 GGTGATGCCGGTTTACGCGACGATCGCGCTGATCATGTTTATGTCGTCGATGGGCTGCCACTGCTGAATGGGTT
 CGTCGGCGAAATGCCGATCATGATGGCGACGTTTATGGTCAACAAGACCTGGGCGATCTGGGCGATTCCGGGCGT
 AGTACTTGCCGCGGCTTATCTGCTGTGGCTCTACCAGCGGGTGTCTTCCGGGACG
 CGTCGCCGTAGAAGCTCTTCTGGTAGAGCCAGAGCAGGTACCACGCGCCCAGATTACGCCCAGGGCTGCGAGGA
 TGCCGTAGATCATCTTCGCTGGAAGGCGCGCTCAGGACGAGGAACCTGCCACGAAGCCGTTGAGCAGCGGCA
 GTCCGGCGGAGGCCAGCGTGATCACCAGGAAGAAGGTGGCGAACACCGGCATCGGCGTGCGGAGGCCGCGCTACT
 GCTTGATCTCGTAAGTGTGGCGGCGTTCTGTAATCATGCCAGCAGCATGAACAGGCCGCGGTCGAGACGCCGT
 GGTTACGATCTGGTACACCGAACCGCTCACGCCATGGCGGTGAAGCTGAAGATGCCAACACGCAGAAGCCAA
 GGTGGCTCACCGACGAGTAGGCGACCAGCTTCTTATGTTTGGCTGGACGAGGGCGACAAGCGCGCCGTAGATGA
 TCCCAATGATTGCCAGCGCGATGATCCAGGGTGCCTTGCGCCGCGCTGCTTCCGGGAACAGCCGAGGTTAAAGC
 GCAGCAGGCCGTAGGTGCCCATCTTGAGGAGCACGCCGGCGAGCAGCACGAACCCGCGGTGGTGTCTTCTACGT
 GCGCGTCGGGACCCATGTATGGAACGGAAGAGTGGGACTTTTACC CGGAAGGCGATGAAGAAGCCGAGGAAGA
 GCCACTGCTGCGCGTGCGAGAAGCCAGAGACGTTTCCGAGTTGGATCGCTGCTGGATTTCAACGAAGTCGAATG
 AGCCAGTCTTGAATACAGCCAGAGCATCGCGGCCAGCATGAACACCGACGCGACCATGGTGTACATGAAGAACT
 TCACCGCGGCGTAGATGCGGCGCTCATGGCCGTAAATGCCGATGAGCAGCGCCATCGGGATCAGCGTGGCTTCCC
 AGAAGAAGTAGAACTGGAAGAGATCGAGCGAGACGAAGACGCCGAGCATCGCCGTCTCAAGGATGAGCATGAGGA
 TGAAGAATTCCTTACGCGTTCCTTGATCGAAGTCCAGGAGATGAGCACGCTCAGCGGAACGAGGAACGTGGTCA
 GCACTACCAGCCAGATCGAAATGCCGTGCGCGCCGAGGTGATAG
 1600170 1600199 30 1600214 1600243 30 14 93.3333 0.0000 50 46.6667 53.3333
 46.4286 53.5714 0.0000 3200413 3200413 TTCACCACAGAGACGCAGAGGTCACAGAGA
 TCTCTGTGACCTCTGTGCCTCTGTGGTGAA
 1849826 1849868 43 1849939 1849980 42 70 81.3953 2.3256 44 54.1176 45.8824
 57.1429 42.8571 0.0000 3699807 3699806
 GAAACCTTTACCCACGAAAGGCACAAAGGGGCACGAAGGATT
 AACCTTTGTGTACTTTGTGTCTTCGTGGTGAAAGCTTTTC
 1849787 1849819 33 1849947 1849979 33 127 84.8485 0.0000 41 57.5758 42.4242
 57.1429 42.8571 0.0000 3699766 3699766 AAAAGAATTAACCACGAAGGACACCAAGTGCAC
 GTGTACTTTGTGTCTTCGTGGTGAAAGCTTTT
 1900181 1900253 73 1900721 1900793 73 467 80.0000 5.3333 67 38.3562 61.6438
 38.3333 61.6667 0.0000 3800974 3800974
 CGCTTTGTATCAGGGCAGCGCTTCAGCGCTGCCGAAGTCCGCCCAATATTTTCGGGGGCTTTAGCCCCCTGC
 GCAGGGGCTGAAGCCCCAAACCTTTTCAGTAGCTCTCTCCGGCAGCGCTGAAGCGCTGCCCTGATACAAACCG
 1900146 1900174 29 1900760 1900788 29 585 96.5517 0.0000 53 34.4828 65.5172
 35.7143 64.2857 0.0000 3800934 3800934 TGTATCAGGCCAGCGCTTCAGCGCTGCCG
 CGGCAGCGCTGAAGCGCTGCCCTGATACA
 1900033 1900061 29 1900760 1900788 29 698 96.5517 0.0000 53 34.4828 65.5172
 35.7143 64.2857 0.0000 3800821 3800821 TGTATCAGGCCAGCGCTTCAGCGCTGCCG

CGGCAGCGCTGAAGCGCTGCCCTGATACA
 1899993 1900021 29 1900760 1900788 29 738 93.1034 0.0000 48 32.7586 67.2414
 33.3333 66.6667 0.0000 3800781 3800781 TGTATCAGGCCAGCGCTTCAGCGCCGCCG
 CGGCAGCGCTGAAGCGCTGCCCTGATACA
 1900721 1900788 68 1900991 1901058 68 202 82.6087 2.8986 74 37.5000 62.5000
 35.0877 64.9123 0.0000 3801779 3801779
 GCAGGGGCTGAAGCCCCAAACCTTTCGAGTAGCTCTCTCCGGCAGCGCTGAAGCGCTGCCCTGATACA
 TGTATCAGGGCAGCGCTTCAGCGCTGCCGAAGTCCCGCCCCTGATATTTTCGGGGCTTCAGCCCCTGC
 1900721 1900796 76 1901140 1901215 76 343 80.5195 2.5974 75 37.5000 62.5000
 35.4839 64.5161 0.0000 3801936 3801936
 GCAGGGGCTGAAGCCCCAAACCTTTCGAGTAGCTCTCTCCGGCAGCGCTGAAGCGCTGCCCTGATACAAACCGTC
 T
 AGCCGCTCTGTATCAGGGCAGCGCTTCAGCGCTGCCGAAGTCCCGCCCCTGATATTTTCGGGGCTTCAGCCCCTG
 C
 1900721 1900796 76 1901277 1901353 77 480 78.4810 6.3291 63 36.6013 63.3987
 35.4839 64.5161 0.0000 3802073 3802073
 GCAGGGGCTGAAGCCCCAAACCTTTCGAGTAGCTCTCTCCGGCAGCGCTGAAGCGCTGCCCTGATACAAACCGTC
 T
 AGCCGCTCTGTATCAGGCCAGCGCTTCAGCGCTGCCGAGCTGCCGCCCTGATATTTTCGGGGCTTCAGCCCCCT
 GC
 1900721 1900788 68 1901421 1901487 67 632 80.0000 7.1429 60 39.2593 60.7407
 37.5000 62.5000 0.0000 3802209 3802209
 GCAGGGGCTGAAGCCCCAAACCTTTCGAGTAGCTCTCTCCGGCAGCGCTGAAGCGCTGCCCTGATACA
 TGTATCAGGGCAGCGCTTCAGCGCTGCCGAAGTCCACCCCTGATATTTTCGGGGCTTCAACCCTGC
 1900721 1900788 68 1901555 1901622 68 766 81.1594 2.8986 69 38.2353 61.7647
 35.7143 64.2857 0.0000 3802343 3802343
 GCAGGGGCTGAAGCCCCAAACCTTTCGAGTAGCTCTCTCCGGCAGCGCTGAAGCGCTGCCCTGATACA
 TGTATCAGGGCAGCGCTTCAGCGCTGCCGAAGTCCCGCCCCTGATATTTTCGGGGCTTTAGCCCCTGC
 2559712 2559752 41 2559844 2559884 41 91 85.3659 0.0000 52 51.2195 48.7805
 51.4286 48.5714 0.0000 5119596 5119596
 CAAAATCTTTAAACACGAAGGACACGAAGGGCGCCCGAAGG
 CCTTCGTGAGACCTTCGTTTCCTTCGTGTTGAAAGCTTTTG
 2698180 2698205 26 2698262 2698287 26 56 92.3077 0.0000 42 57.6923 42.3077
 58.3333 41.6667 0.0000 5396467 5396467 GGAGTTAGGTTGTTAGGAGTTAGGTT
 AACCTAATTCCTAACCAACCTAACGCC
 2922271 2922312 42 2922329 2922370 42 16 83.3333 0.0000 49 60.7143 39.2857
 60.0000 40.0000 0.0000 5844641 5844641
 AAAATCTTTAAACACGAAGGGCACGAAGTACACGAAGGTTTT
 AAGACCTTCGTTTCCTTCGTGACCTTCGTGTTAAACGAATTT
 3085849 3085896 48 3086408 3086455 48 511 77.0833 0.0000 41 43.7500 56.2500
 45.9459 54.0541 0.0000 6172304 6172304
 GGTCGTTGCAATCACCTTAGTCAGCGACGCCATGTGCAAGATCGTGTC
 GAGACGATCTTCGATCTGGCATCGGTAACCAAGGCCGTTGCGACGACC
 3103548 3103625 78 3103628 3103704 77 2 77.5000 6.2500 60 52.9032 47.0968
 53.2258 46.7742 0.0000 6207253 6207251
 CTAACCTCAAAACCTTTCACCACGAAGGACACGAAGTAAACGAAGGTGGCTTTGTGTTGCAAGATCGACGGCTTC

GGT

ACCGAAATCGTCAACCTAACCAACACAAACCCTTCGTGAACCTTCGTGTCCTTCGTGGTTAAGGCCTTTAGCAGTT
AG

3661358 3661397 40 3662002 3662042 41 604 82.9268 2.4390 45 54.3210 45.6790
52.9412 47.0588 0.0000 7323399 7323399

GCCCGATTACTTGTGGTAAGGGGATTTACCAGCAGTAAAT

ATTTGCTGCTGGTAACAGCCCTTACCACTAGTAAATCGTGC

3949705 3949751 47 3949833 3949878 46 81 87.2340 2.1277 62 55.9140 44.0860
56.0976 43.9024 0.0000 7899584 7899583

AAAGCAAAATCTTTAACCACGAAGGGCACGAAGGGAACACGAAGGAT

ATCCTTCGTGGTACTTCGTGTCCTTCGTGGTGAAGTTTTTGCTTT

4176289 4176320 32 4176339 4176370 32 18 87.5000 0.0000 44 46.8750 53.1250
46.4286 53.5714 0.0000 8352659 8352659 TTAACCACGAAGGACACAAAGGGCCACGAAGG

CCTTCGTGAACCTTGTGGCCTTCGTGGTTAA

4965518 4965555 38 4965573 4965611 39 17 84.6154 2.5641 46 55.8442 44.1558
57.5758 42.4242 0.0000 9931128 9931129

AAATCTTTAGACACGAAGGACACTAAGTTCACGAAGGA

TCCTTCGTGAGACTTCGTGTCCTTCGTGGTTAAGGACTT

5050687 5053704 3018 5062886 5065828 2943 9181 94.8133 3.0723 5083 41.0502
58.9498 41.0105 58.9895 0.0000 10116590 10116590

TCGACCGTGATGGATAAGAGCGCCATAGCACATACTCCTCTTTTGCCAAAGGCCTTTACTGATGAGTAACGATTT

GGGTAGGAGCGGTGGGCGCGGGGCGAGTGGCTGGGAGGGGTAAAGATCCGCTTAACCGATGCGGACCTATACTA

TTGGATCATCCCCACCCCGGCGGGGGCGGAGTTGTCGCCAAGTTGCCATTCCCTCGTGGAAGGAGCTCAAAAAA

GGCCTGCTACCCCCCTTTTTCGTGACCCCACTTTTGGGGCCCGTGCGATGTCTTTGGAAGCAACATTTATCTCT

TATTTTTTAGCCTACTCGGGCTGCCAAGCCCCGAGCCAGGTTGAGGGGCACCCAGCCCCGCGGCGGGCGAGAAGAG

GGTTGATCTGCGATATCGAAATGCTTGCCCGCATGCCGCCGAAGTGGAAAGAGCTGCTGCTGCAAGGCTTCTCGA

CGCGTAAAGTCGTGCGCGCTGGCGGGCGGAAACCGCCCAAGAACTGCCTGACGAGATCGGTTTCGAGAAGCTC

TCCGCAGTCGGCGGGGAGCTTTCTTTTGTGGGAAGAACACGGCGTTGAGGAGAGGCCGGGGGCGAGCAAGACGG

GAGCAGGCATACCCGCCCTTACAAATATTTACAGGGCAAAACGTCTTCTTAACCGAGGAATTGCTATGTACCCAAA

ATCTCCGATAGTTGTGCGCCCAAGTTGCCATTCCCCCGTGGCGAAGGGTCGAAAAAGCCTCGCTACCCCCCTTTT

TCCTCGACCCCAATTTGAGGCCCGGCTCAGGCAGTTTTTTCCTGAGGTTCAAACGGATTTGCACACCAAGTATCTG

TAAACGGTCACTCTGACGCCTAATGGGAGCCCAGGTTGAGGGGCGTCCAGCCCCGGCGGGCGGAGAGAAAGA

GATACTGATCCACCTACCTGCTTCAACCTAAATAAAGCCGACTCACCACCGAGGTGAGGGAGCTTTATTACTCA

CAGCACAACCCCCGGCTCCCAACACCCACACTCTCCGTGGAACGCCGACGACCAACCAGGAGACGCTCGGCTCG

ATACCGAACCCTGGGCACGAAGCCCACGCTGTTGGTAACCGTGGACGACCAACCCAGAAGACTCCGGCTCCGAGC

CTCGCCCCGAACAACAACTCTTCCGAGCCCACGCTCCTGGTCAAACCTCGACGACAACCCGGAGGCGTTCCGGCTC

GAGCTTCGGCCCCGCACGAGGATTTTTACTCGCAGCGAACTACGAACGCTGCTCCCGAGGTCCACGGTCTTGGT

TAACTGCGGACGACAACTCGGAGACGTTGGGCTTCGATCCTCGCCCCGCGGAGGATTTTTACTTTCCGCAGCG

AACTACTAGGTGGAAGCGAAAAAAGGGTCTCTTGATGGAAGTTGGGCGACAGCGAGCGACGAAGGTGGCCTGC

ATTTCTGGACGGCATCGTCGATCGACACACCGGATCCTTCGTTGAGGATCTCGACGCCGAAGATCTTCATCGTG

GTGGCGGAACCCTGTTGTTTCGTACCGCAGAGGGTGATGTGGAAGGGCAGGATCTGGTCGGAGTACCAGGGCTTG

GCCTGCTCCTTGAAGCCGGAGACGAGGGTGTTCTGGATACCGGAGTGGGAAGCCTTTGGTAGATATCGCTCCTGG

TTCTAATCTCATTCTTGACATACCGCCGACATATCGGCATATACTTACGGCCCCGGCGATTATCCCGGCCGAACCA

GGAGAGCTCTATGAAACAGAATAGATGTGCGCGCCATCTCCGTCTCGCCCGTCCCGTTACAGATCAACCAGCA

GGTGGGGCGTGATGCTCTGCCATCCAGGCCCTGCAGCTTGAGGCCTTGATCGAACTCCAGAAGCCGGCGCCGAG

CCTGCCACGCCTCCAGTTCCTGAACCGCATGATCGACCTGGCCAAGGACGAGGTTGTCAGCCTGGAGAAGTTCTA

ATGAAACGCGCCGCGATTTACTTGCGGGTGTCCACGCTCGACCAACACCCCGAAACCCAACTCTACGACCTTGAA
GAACTGGCGCGACAGCGCCAGTTCGAGATCGTCGCAACCTATGAAGACCGGATCAGCGGCGCGAAGGCGCGCCG
CCGGGGCTCGATCAGATGATGAAGGACGCCCCTGCCACCACTTCGATGTGGTGGTCTGTGGGCATGTGACCGC
CTGGCGCGATCGGTGAAGCACTTCTTAGAAACGTTAGATGAGTTGACTCATCTCAACATCGAGTTCTGAAGCTTC
CGCGAACAGCTCGACACCGCGCGCGTTGGGCCGCGCCGTCTGACCATCATCAGTGTGTGCGCCGAAGTGGAA
CGCTCTCTCATCATCGAGCGGGTGCCTGCCGGCATGCGCCGCGCAAAGTTAGAGGGCACCCACATCGGCCGCAA
CCGCTCGATATCGATCGCCAAGCGGTGCTGCGGCACCGTGCTCACGGTCAGTCGCTGACCCAGATCGCCAAGACC
TTCAGCATCGGAAGAGCGACTGTCTCCCGCATCATTAAAGAAGAAGCCAGGCCCTTCTCTGGCTGTTGGCGAA
GCGTAATGGTGAACGTCTCCGATACGCTGCGCGACTATCGTCGCTGCCTGCGGAAACGCCGCTATCAGACGCGGG
AAGCGGCCGAGCAGAAGGCACGCAAAGCGAGCACACGAACCGGTGAGGTGATCCACTCCTACCACTGCGGTCTAT
GCGGACTGTGGCATATCGGAAAGCAACGTCCCCGAAGGGGCAAACGATCGAAGTTCGGGTGGCCGGTTGTGCT
GGTTGTGCGGCGATCCGATCCCCGTGAGCGTCTGGAGAAATCGTCTCATGCACATCCGACGTTTACCTGCTCGC
CTCGCTGCAGCAAGCTCAGAAAAAGGTTCTCGCCAAGCAGCGTCGCATGAAGAAGGCACGGGAAGACCTTGGCC
CAAAAGGTCTTCCGGAACCTCGTCCTAACTTAAACAAAACAGGGCGCCGGAACCGCGCCCTTCACTAGCCAC
AAGGAGTGGGTTTCCGGC
GCCCCAAACCCACTCCTTGTTGGGCTAGTGAAGGGCGCCGTTTCCGGCGCCCTGTTTTGTTTAAGTTAGGACGAGG
TTCCGGAAGACCTTTTGGGCCAAGGTCTTCCCGTGCCTTCTTCATGCGACGCTGCTTGGCGAGGAACCTTTTTC
TGAGCTTGCTGCAGCGAGGCGAGCAGGTAAACGTGGATGTGCATGAGACGATTTCTCCAGACGCTCACGGGGGA
TCGGATCGCCGCACAACCAGCACAACCGGCCACCCGGAACCTTCGATCGTTTCCCCCTTCGGGGGACGTTGCTTTC
CGATATGCCACAGTCCGCATAGACCGCAGTGGTAGGAGTGGATCACCTCACCGGTTCTGTGTGCTCGCTTTCGCTG
CCTTCTGCTCGGCCGCTTCCCGCTCTGATAGCGGCGTTTCCGCAGGCAGCGACGATAGTCGCGCAGCGTATCGG
AGACGTTACCAATTACGTTTCGCCAACAGCCAGAGGAAGGGCCTGGCGTTCTTCTTAAATGATGCGGGAGACAGT
CGCTCTTCCGATGCTGAAGGTCTTGGCGATCTGGGTGAGCGACTGACCGTGAGCACGGTGCCGACGACCGCTTG
GCGATCGATATCGAGCGGTTTGGGCCGATGTGGGTGCCCTCTAACTTTGCGCGGCGCATGCCGGCACGCACCCG
CTCGATGATGAGAGAGCGTTCCAGTTCGGCGACAACACTGATGATGGTCACGACGGCGCGGCCAACGCGCCGGC
GGTGTGAGCTGTTGCGGGAAGCTTACGAACTCGATGTTGAGATGAGTCAACTCATCTAACGTTTCTAAGAAGTG
CTTACCGATCGCGCCAGGCGGTACATGCCACACGACCACCACATCGAAGTGGTGGCGACGGGCGTCTTCAT
CATCTGATCGAGCCCCGGGCGCGCCTTCGCGCCGCTGATCCGGTCTTCATAGGTTGCGACGATCTCGAACTG
GCGCTGTGCGCCAGTTCTTCAAGGTCTGAGATTGGGTTTCCGGGTGTTGGTCGAGCGTGGACACCCGCAAGTA
AATCGCGGCGGTTTCATTAGAACTTCTCCAGGCTGACAACCTCGTCCTTGGCCAGGTCGATCATGCGGTTACGG
AACTGGAGGCGTGGCAGGCTCGGCGCCGGCTTCTGGAGTTCGATCAAGGCCTCAAGCTGCAGGGCCTGGATGGCA
GAGCATACACGCCCCACCTGCTGGTTGATCTGTGAACGGGACGGGCGAGGACGGAGATGGCGGCGACATCTATTC
TGTTTCATAGAGCTCTCCTGGTTTCGGCCGGGATAATCGCCGGGCGTAAGTATATGCCGATATGTGCGCGGTATG
TCAAGAATGAGATTAGAACCAGGAGCGATATCTACCAAAGGCTTCCCACTCCGGTATCCAGAACCCCTCGTCTC
CGGCTTCAAGGAGCAGGCCAAGCCCTGGTACTCCGACCAGATCCTGCCCTTCGACATCACCTCTGCGGTACGAA
CGAACAGGGTTCCGCCACCACGATGAAGATCTTCGGCGTCGAGATCCTCAACGAAGGATCCGGTGTGTCGATCGA
CGATGCCGTCCAGGAAATGCAGGCCACCTTCGTCGCTCGCTGTGCCCCAACTTCCATGCAAGAGACCTTTTTTC
GCTTCCGACCTAGTAGTTCTGCTGCGGAAAGTAAAAATCCTCGCGCGGGGGCGAGGATCGAAGCCGAACGTCTCCG
AGTTGTCTGTCGCGAGTTAAACCAAGACCGTGGGACCTCGGGAGCAGCGTTCTGATGTTCTGCTGCGAGTAAAAATCCT
CGTGCGGGGGCCGAAGCTCGAGCCGAACGCCTCCGGGTTGTCTGTCGAGGTTTGACCAGGAGCGTGGGCTCGGAAGA
GTTGTTGTTTCGGGGCGAGGCTCGGAGCCGGAGTCTTCTGGGTTGGTCTCCACGGTTACCAACAGCGTGGGCTTC
GTGCCCAGGGTTCCGGTATCGAGCCGAGCGTCTCCTGGTTGGTCTGCGCGTTCACGGAGAGTGTGGGGTGTGG
GAGCCGGGGGTTGTGCTGTGAGTAATAAAGCTCCCTCACCTCGGTGGTGAGTCGGCTTTATTTAGGTTGAAGCAG
GTAGGTGGGATCAGTATCTTTTCTCTCGCCCCGCCGGGGGCTGGACGCCCTCAACCTGGGCTCCCATTAGGC
GTCAGAGTGACCGTTTTACAGATACTGGTGTGAAATCCGTTTGAACCTCAGGAAAAAACTGCCTGAGCCGGCCT
CGAAATTGGGGTCGAGGAAAAAGGGGGTAGCGAGGCTTTTTCGACCCTTCGCCACGGGGGAATGGCAACTTGGG

CGACAACTATCCGGAGATTTTGGGTACATAGCAATTCCTCGGTAAAGAAGACGTTTTGCCCTGTAAATATTTGTA
AGGCGGGTATGCCTGCTCCCGTCTTCGCTGCCCCGGCCTCTCCTGAACGCCGTGTTCTTCCCAGCAAAAGAAAG
CTCCCCGCCGACTGCGGAGAGCTTCTCGAAACCGATCTCGTCAGGCAGTTAGGTCCGCAATGACTTAGGAGATGG
GGCTGGTACTATCATCCCTCTTCTCGCCCGCCGCGGGGCTGGACGCCCTCAACCTGGCTCCCATAGGCGTCA
GAGTGGGCTAAAAGTAGGAGGCAAATGTTGTCTCCCGAGGCATTGCCCGGCCTGAGAATGGGGGTAGAGGAAAA
AGGGGGATAGCGAGGTTTTTTGACCCCTTCGCCACGAGGGAATGGCAACTTGGGCGACAACTCCACCCCCAGCCG
GGGAGGGGATGATCCAATAGTATAGAGTCGCATCGATTAAGCGGACCTTTGCCCTACCATCCACTCTCTCCCC
GTCTGACGCGCATACCCAAATCGTTACTCATCAGTACACGCCTTTGGCTCACATACGTCGCTGTGCTAGGTGCGC
TCTCACCATTGCGGTCTGA

5305712 5305747 36 5305771 5305806 36 23 83.7838 5.4054 40 55.5556 44.4444
54.8387 45.1613 0.0000 10611518 10611518

TCTTTAAACACGAAGGACACTGAGGAAGCACGAAGG CCTTCGTGATGCCTTCATGTCTTTCGTGTTTTAAGA
5484048 5484097 50 5484300 5484350 51 202 82.6923 5.7692 53 50.4950 49.5050
51.1628 48.8372 0.0000 10968397 10968397

TTCACCACGAAGGACACGAAGGGCACGAAGTTTTTAGGAATTAGGTAT
ATCCCCAACTTCAAATACCTTCGTGGGCCTTCGTGTCCTCCGTGGTTGAA

Table S5. Repeats

Program	No. in Ellin 6076 (size range)	No. in Ellin 345 (size range)
repeat-match	63,764 (20 – 3027)	7294 (20 – 2525)
exact-tandems	31 (3 – 261)	19 (7 – 100)
Tandem Repeats Finder	302 (3 – 417)	154 (6 – 279)
Inverted Repeats Finder	120 (23 – 4436)	34 (21 – 3018)

Table S6. Ellin6076 mobile elements

Type	Locus tag	Domains from NCBI Conserved Domain Search	Identity group
phage integrase family protein	Acid_0347	INT_XerDC	none
phage integrase family protein	Acid_0629	COG4974 - Site-specific recombinase XerD INT_Int1	5
phage integrase family protein	Acid_0630	INT_pAE1 Phage integrase, N-terminal SAM-like domain. COG4974 - Site-specific recombinase XerD	9
phage integrase family protein	Acid_0632	INT_SG5 COG4974 - Site-specific recombinase XerD	10
phage integrase family protein	Acid_0633	INT_pAE1 COG4974 - Site-specific recombinase XerD	none
transposase IS3/IS911 family protein	Acid_0634	COG2963	1
integrase catalytic region	Acid_0635	pfam00665 rve	1
Transposase, IS204/IS1001/IS1096/IS1165 family	Acid_0644	COG3464	none
integrase catalytic region	Acid_0719	pfam00665 rve HTH_7	6
integrase catalytic region	Acid_0819	pfam00665 rve pfam08281 Sigma70_r4_2	6
integrase catalytic region	Acid_0942	pfam00665 rve	2
transposase IS3/IS911 family	Acid_0943	pfam01527 transposase_8	2
integrase catalytic region	Acid_1129	pfam00665 rve HTH_7	6
transposase	Acid_1319	pfam01548 Transposase_9 pfam02371	7

IS116/IS110/IS902 family		transposase_IS116/IS110/IS902	
integrase catalytic region	Acid_1329	pfam00665 rve HTH_7	6
transposase IS3/IS911 family	Acid_1503	Pfam01527 transposase_8	3
integrase catalytic region	Acid_1504	pfam00665 rve	3
integrase catalytic region	Acid_1623	pfam00665 rve	none
transposase IS3/IS911 family	Acid_1624	COG2963	1
integrase catalytic region	Acid_1625	pfam00665 rve	1
putative transposase protein Y4bF	Acid_1944	no domains defined	8
putative transposase protein Y4bF	Acid_1955	100% identical to Acid_1944 no domains defined	8
integrase catalytic region	Acid_2150	pfam00665 rve HTH_7	6
transposase IS116/IS110/IS902 family	Acid_2198	Pfam02371 Transposase IS116/IS110/IS902 family	11
integrase catalytic region	Acid_2230	pfam00665 rve HTH_7	6
integrase catalytic region	Acid_2273	pfam00665 rve	2
transposase IS3/IS911 family	Acid_2274	Pfam01527 transposase_8	2
integrase catalytic region	Acid_2459	pfam00665 rve	2
transposase IS3/IS911 family	Acid_2460	Pfam01527 transposase_8	2
phage integrase family	Acid_2520	INT_pAE1	none
phage integrase family	Acid_2521	INT_SG4 COG4974 - Site-specific recombinase XerD	none
phage integrase family	Acid_2522	INT_SG5 COG4974 - Site-specific recombinase XerD	none
transposase IS66	Acid_2528	pfam03050	none
transposase IS3/IS911 family	Acid_2592	Pfam01527 transposase_8	3
integrase catalytic region	Acid_2593	pfam00665	3

transposase IS116/IS110/IS902 family	Acid_2663	rve pfam01548 Transposase_9 pfam02371 Transposase_20 transposase_IS116/IS110/IS902	7
transposase IS116/IS110/IS902 family	Acid_2712	pfam01548 Transposase_9 pfam02371 Transposase_20 transposase_IS116/IS110/IS902	7
phage integrase family protein	Acid_2776	INT_Int1 COG4974 - Site-specific recombinase XerD	4
putative transposase	Acid_2778	pfam04986 - Transposase_32 IPR007069 Transposase, IS801/IS1294	4
transposase IS116/IS110/IS902 family	Acid_2878	pfam01548 Transposase_9 pfam02371 Transposase_20 transposase_IS116/IS110/IS902	7
phage integrase family protein	Acid_2897	INT_phiLC3_C Pfam01935 DUF87 Pfam00589?	none
transposase IS66	Acid_2942	pfam03050 Transposase IS66 family	12
phage integrase family protein	Acid_2943	INT_SG5 COG4974 - Site-specific recombinase XerD	none
phage integrase family protein	Acid_2945	INT_pAE1 COG4974 - Site-specific recombinase XerD	none
transposase IS66	Acid_2946	no conserved domains	none
transposase IS3/IS911 family	Acid_3155	Pfam01527 transposase_8	2
integrase catalytic region	Acid_3156	pfam00665 rve	2
transposase IS116/IS110/IS902 family	Acid_3180	pfam02371 Transposase_20 transposase_IS116/IS110/IS902	13
putative transposase protein Y4bF	Acid_3260	no domains defined 100% identical to the other Y4bF genes	8
putative transposase	Acid_3308	pfam04986 - Transposase_32	4

phage integrase family protein	Acid_3310	IPR007069 Transposase, IS801/IS1294 INT_pAE1 Phage integrase, N-terminal SAM-like domain. COG4974 - Site-specific recombinase XerD	9
phage integrase family protein	Acid_3311	INT_SG4 COG4974 - Site-specific recombinase XerD	14
phage integrase family protein	Acid_3312	INT_SG5 COG4974 - Site-specific recombinase XerD	10
phage integrase family protein	Acid_3313	INT_Int1 COG4974 - Site-specific recombinase XerD	4
integrase catalytic region	Acid_3574	pfam00665 rve	2
transposase IS3/IS911 family	Acid_3575	Pfam01527 transposase_8	2
integrase catalytic region	Acid_3877	pfam00665 rve HTH_7	6
transposase IS116/IS110/IS902 family	Acid_3943	pfam01548 Transposase_9 pfam02371 Transposase_20 transposase_IS116/IS110/IS902	7
transposase IS116/IS110/IS902 family	Acid_3955	pfam01548 Transposase_9 pfam02371 Transposase_20 transposase_IS116/IS110/IS902	7
transposase IS116/IS110/IS902 family	Acid_4031	pfam01548 Transposase_9 pfam02371 Transposase_20 transposase_IS116/IS110/IS902	7
transposase IS116/IS110/IS902 family	Acid_4465	pfam01548 Transposase_9 pfam02371 Transposase_20 transposase_IS116/IS110/IS902	13
integrase catalytic region	Acid_4625	pfam00665 rve HTH_7	6
phage integrase family	Acid_4695	INT_P4	none
integrase catalytic region	Acid_4733	pfam00665 rve	3
transposase IS3/IS911 family	Acid_4734	Pfam01527 transposase_8	3

transposase	Acid_4746	COG5659 FOG: Transposase	none
putative transposase	Acid_4751	pfam04986 Putative transposase	4
phage integrase family	Acid_4752	INT_Int1	4
		COG4974 - Site-specific recombinase XerD	
transposase IS3/IS911 family	Acid_5157	COG2963 Transposase and inactivated derivatives	1
		pfam01527 - Transposase_8	
integrase catalytic region	Acid_5158	pfam00665 rve	1
integrase catalytic region	Acid_5161	pfam00665 rve	2
transposase IS3/IS911 family	Acid_5162	pfam01527 - Transposase_8	2
integrase catalytic region	Acid_5236	pfam00665 rve	6
		HTH_7	
transposase	Acid_5320	pfam01548 Transposase_9 pfam02371	7
IS116/IS110/IS902 family		Transposase_20	
		transposase_IS116/IS110/IS902	
transposase	Acid_5341	pfam02371 Transposase_20	11
IS116/IS110/IS902 family		transposase_IS116/IS110/IS902	
integrase catalytic region	Acid_5347	pfam00665 rve	6
		HTH_7	
transposase	Acid_5398	pfam01548 Transposase_9 pfam02371	7
IS116/IS110/IS902 family		Transposase_20	
		transposase_IS116/IS110/IS902	
putative transposase protein Y4bF	Acid_5588	no domains defined	8
transposase	Acid_5784	pfam01548 Transposase_9 pfam02371	7
IS116/IS110/IS902 family		Transposase_20	
		transposase_IS116/IS110/IS902	
transposase	Acid_5830	pfam01548 Transposase_9 pfam02371	7
IS116/IS110/IS902 family		Transposase_20	
		transposase_IS116/IS110/IS902	
phage integrase family	Acid_5994	INT_phiLC3_C	none
		COG4974 - Site-specific recombinase XerD	

transposase IS66	Acid_6003	pfam03050 - Transposase_25 COG2251 - Predicted nuclease (RecB family)	none
integrase catalytic region	Acid_6059	pfam00665 rve	15
integrase catalytic region	Acid_6069	pfam00665 rve	2
transposase IS3/IS911 family	Acid_6070	pfam01527 - Transposase_8	2
transposase	Acid_6101	no domains, fragment?	5
phage integrase family protein	Acid_6102	INT_Int1 COG4974 - Site-specific recombinase XerD	5
putative transposase	Acid_6119	pfam04986 Putative transposase	none
putative transposase	Acid_6124	no domains	12
transposase	Acid_6356	pfam01548 Transposase_9 pfam02371	7
IS116/IS110/IS902 family		Transposase_20 transposase_IS116/IS110/IS902	
transposase	Acid_6476	pfam01548 Transposase_9 pfam02371	7
IS116/IS110/IS902 family		Transposase_20 transposase_IS116/IS110/IS902	
integrase catalytic region	Acid_6521	pfam00665 rve	6
transposase	Acid_6688	pfam01548 Transposase_9 pfam02371	7
IS116/IS110/IS902 family		Transposase_20 transposase_IS116/IS110/IS902	
transposase	Acid_6781	no domains, fragment?	none
phage integrase family	Acid_6791	INT_XerDC COG4974 - Site-specific recombinase XerD	none
phage integrase family	Acid_6799	INT_SG5 COG4974 - Site-specific recombinase XerD	none
phage integrase family	Acid_6800	INT_SG4 COG4974 - Site-specific recombinase XerD	none
integrase catalytic region	Acid_6803	COG3316	16
integrase catalytic region	Acid_6807	pfam00665 rve	none
integrase catalytic region	Acid_6822	pfam00665 rve	1
transposase IS3/IS911 family	Acid_6823	pfam01527 - Transposase_8	1

transposase	Acid_6843	no domains	none
putative transposase	Acid_6844	no domains	none
transposase IS3/IS911 family	Acid_6861	pfam01527 - Transposase_8	1
integrase catalytic region	Acid_6862	pfam00665 rve	1
transposase	Acid_6867	no domains	none
transposase	Acid_6898	COG3464 Transposase and inactivated derivatives	none
IS204/IS1001/IS1096/IS1165 family		pfam01610 - Transposase_12	
transposase	Acid_7041	pfam01548 Transposase_9 pfam02371	7
IS116/IS110/IS902 family		Transposase_20	
integrase catalytic region	Acid_7104	transposase_IS116/IS110/IS902	
		pfam00665 rve	6
		HTH_7	
transposase-like	Acid_7107	KEGG: mxa:MXAN_0516 transposase, IS630 family	none
		pfam01371:Trp repressor protein. This protein binds to tryptophan and represses transcription of the Trp operon.	
integrase catalytic region	Acid_7211	pfam00665 rve	6
		HTH_7	
transposase	Acid_7233	pfam01548 Transposase_9 pfam02371	7
IS116/IS110/IS902 family		Transposase_20	
		transposase_IS116/IS110/IS902	
transposase IS3/IS911 family	Acid_7255	pfam01527 - Transposase_8	3
integrase catalytic region	Acid_7256	pfam00665 rve	3
integrase catalytic region	Acid_7608	pfam00665 rve	16
integrase catalytic region	Acid_7618	pfam00665 rve	15
integrase catalytic region	Acid_7626	pfam00665 rve	16
phage integrase family protein	Acid_7640	DNA-BRE-C	5
phage integrase family protein	Acid_7641	DNA-BRE-C, Site-specific recombinase XerD	10

phage integrase family protein	Acid_7642	Int_SG4, DNA-BRE-C, site-specific tyrosine recombinase XerC	14
phage integrase family protein	Acid_7643	pAE1, DNA-BRE-C, XerC	9
putative transposase	Acid_7645	pfam04986 Putative transposase	5
integrase catalytic region	Acid_7942	pfam01527 - Transposase_8 pfam00665 rve	3

Table S7. Candidate genomic islands in the Ellin 6076 genome

Region	size/coords	tRNA location	mobile elements/phage/plasmid genes in the region	GC content	dinucleotide bias
I. Acid_0037 – Acid_0155	57495-199752	Acid_R0002 on + strand between Acid_0139 and Acid_0143	no phage Tail Collar domain in Acid_0149 Acid_0150 Acid_0151	variable	variable
II. Acid_0208 – Acid_0379	273165-480673	Acid_R0003 on – strand between Acid_0243 and Acid_0230	phage integrase family protein Acid_0347	variable	variable
III. Acid_0607 – Acid_0665 Good candidate	745512 - 832975	Acid_R0007 on + strand between Acid_0604 and Acid_0608	phage integrase family proteins Acid_0629 Acid_0630 Acid_0632 Acid_0633 Acid_0635 transposase IS3/IS911 family Acid_0634 transposase IS204/IS1001/IS1096/IS1165 family Acid_0644 plasmid maintenance system	mostly below average	yes

			killer Acid_0688 plasmid maintenance system antidote Acid_0689 plasmid pRiA4b ORF-3 family protein Acid_0696		
IV. Acid_0819 – Acid_1054	1017689- 1194647	Acid_R0009 on + strand between Acid_0970 and Acid_0974 Acid_R0011 and Acid_R0012 in an rRNA operon between Acid_1004 and Acid_1007	integrase catalytic region Acid_0819 integrase catalytic region Acid_0942 transposase IS3/IS911 family Acid_0943 plasmid stabilization system Acid_1015	variable	yes
V. Acid_1110 – Acid_1160	1365753- 1431894	Acid_R0015 on – strand between Acid_1158 and	integrase catalytic region Acid_1129	variable	variable

Acid_1161					
VI.	1500998-	Acid_R0016	transposase IS116/IS110/IS902	variable	variable
Acid_1216 –	1858185	on – strand	family		
Acid-1514		between	Acid_1319		
		Acid_1369	integrase catalytic region		
		and Acid-1370	Acid_1329		
			transposase IS3/IS911 family		
		Acid_R0017	Acid_1503		
		on + strand	integrase catalytic region		
		between	Acid_1504		
		Acid_1372			
		and			
		Acid_1373			
		Acid_R0018			
		on + strand			
		between			
		Acid_1475			
		and			
		Acid_1477			
VII.	1974941-	Acid_R0020	integrase catalytic region	variable	variable
Acid_1602 –	2200453	and	Acid_1623		
Acid_1770		Acid_R0021	Acid_1625		
		on + strand	transposase IS3/IS911 family		
		in an rRNA	Acid_1624		
		operon			
		between	Putative bacteriophage-related		
		Acid_1601	Acid_1678		

		and Acid-1602	plasmid-related protein Acid_1609		
		Acid_R0024 on + strand between Acid_1607 and Acid_1609			
		Acid_R0025 on + strand between Acid_1665 and Acid_1669			
		Acid_R0026 on + strand between Acid_1770 and Acid_1771			
VIII. Acid_1830 – Acid_2150	2282350- 2752528	Acid_R0027 on + strand between Acid_1923 and Acid_1928	putative transposase protein, Y4bF Acid_1944 Acid_1955 integrase catalytic region Acid_2150	variable	variable

		Acid_R0028 on – strand between Acid_2113 and Acid_2114			
IX. Acid_2196 -Acid_3156	2814111- 3999816	Acid_R0029 on – strand between Acid_2346 and Acid_2348 Acid_R0030 on – strand between Acid_2688 and Acid_2691 Acid_R0031 on + strand between Acid_2749 and Acid_2752 Acid_R0032 on + strand between	transposase IS116/IS110/IS902 family Acid_2198 IS3/IS911 family Acid_2274 Acid_2460 IS66 Acid_2528 Acid_2529 IS3/IS911 family Acid_2592 IS116/IS110/IS902 family Acid_2663 Acid_2712 putative transposase Acid_2778 IS66 Acid_2940 Acid_2942 Acid_2946 IS3/IS911 family Acid_3155 Integrase catalytic region	variable	variable

Acid_2829	Acid_2230
and	Acid_2273
Acid_2833	Acid_2459
	phage integrase family
Acid_R0033	Acid_2520
on + strand	Acid_2521
between	Acid_2522
Acid_2895	integrase catalytic region
and	Acid_2593
Acid_2898	phage integrase family
	Acid_2776
	Acid_2943
	Acid_2945
	integrase catalytic region
	Acid_3156
	plasmid maintenance system
	antidote protein
	Acid_2330
	plasmid maintenance system
	killer
	Acid_2331
	phage SPO1 DNA polymerase-
	related protein
	Acid_2350
	phage shock protein A
	Acid_2384

X. Acid_3253 – Acid_4625 a lot of mobile elements, tRNA genes and small areas of dinucleotide bias – may have been an island at one time, but subsequent rearrangements obscured the boundaries?	4117534- 5850429		phage shock protein C Acid_4337 phage shock protein A Acid_4570	variable	variable
XI. Acid_4625 – Acid_4695 Good candidate	5850429- 5944933	Acid_R0043 on – strand between Acid_4694 and Acid_4696	Integrase catalytic region Acid_4625	mostly below average	yes
XII. Acid_4696 – Acid_5994 a lot of mobile elements, tRNA genes and small areas of dinucleotide	5945173- 7519982		bacteriophage N4 receptor Acid_5606	variable	variable

bias – may
have been an
island at one
time, but
subsequent
rearrangements
obscured the
boundaries?

XIII.	7502911- Acid_5980 – Acid_6176	7745505	Acid_R0050 on + strand between Acid_5992 and Acid_5997 Acid_R0051 on + strand between Acid_6067 and Acid_6072	Transposase IS66 Acid_6003 transposase Acid_6101 putative transposase Acid_6119 IS66 Acid_6121 putative transposase Acid_6124 Integrase catalytic region Acid_6059 Acid_6069 phage integrase family Acid_6102	variable	variable
XIV.	Acid_6243 – Acid_6638 hard to find	7825707- 8326556		phage Gp37Gp68 family protein Acid_6594 plamid stabilization system	variable	variable

any boundaries here		Acid_6411			
XV.	8469801-	Acid_R0057	transposase	mostly below average	yes
Acid_6758 –	8733305	on + strand	Acid_6781		
Acid_6949		between	IS3/IS911		
		Acid_6765	Acid_6823		
Good		and			
candidate		Acid_6770	transposase		
			Acid_6843		
			putative transposase		
			Acid_6844		
			IS3/IS911		
			Acid_6861		
			transposase		
			Acid_6867		
			IS204/IS1001/IS1096/IS1165		
			family		
			Acid_6898		
			phage integrase family		
			Acid_6791		
			Acid_6799		
			Acid_6800		
			integrase catalytic region		
			Acid_6803		
			Acid_6807		
			Acid_6822		
			Acid_6862		

XVI.	9059709-	Acid_R0058	integrase catalytic region	variable	variable
Acid_7211 –	9616304	on – strand	Acid_7211		
Acid_7657		between	transposase IS116/IS110/IS902		
		Acid_7657	family		
		and	Acid_7233		
		Acid_7658	transposase IS3/IS911 family		
			Acid_7255		
			integrase catalytic region		
			Acid_7256		
			Acid_7608		
			Acid_7618		
			Acid_7626		
			phage integrase family protein		
			Acid_7640		
			Acid_7641		
			Acid_7642		
			Acid_7643		
			putative transposase		
			Acid_7645		
			plasmid-encoded RepA protein		
			Acid_7603		
			plasmid-related protein		
			Acid_7627		

Table S8. Distribution of genes in COG categories for *Acidobacteria* strains Ellin 6076 and Ellin 345, compared to other large-small genome pairs.

COG Category	6076	345	MLOT	BNC1	MC2 155	MLEP	HCHE	MED92	REU	RSO
Information storage and processing										
A. RNA processing and modification	0	1	0	0	3	1	2	1	1	5
B. Chromatin structure and dynamics	3	2	5	1	1	0	3	1	3	2
J. Translation, ribosomal structure and biogenesis	183	164	198	169	177	123	191	182	207	155
K. Transcription	471	295	569	289	525	74	397	251	604	314
L. Replication, recombination and repair	328	158	294	164	193	65	212	126	192	212
Cellular processes and signaling										
D. Cell cycle, cell division, chromosome partitioning	26	25	34	35	28	21	40	32	37	29
M. Cell wall/membrane/envelope biogenesis	443	257	313	228	170	69	269	168	301	242
N. Cell motility	101	82	67	51	6	6	185	121	113	129
O. Posttranslational modification, protein turnover, chaperones	183	134	205	163	136	62	201	143	219	151
T. Signal transduction mechanisms	421	316	254	135	190	39	474	380	302	212
U. Intracellular trafficking, secretion and vesicular transport	172	108	141	140	24	20	142	84	144	167
V. Defense mechanisms	236	116	67	40	49	12	69	34	65	46
W. Extracellular structures	0	0	1	0	0	0	0	0	0	1
Y. Nuclear structure	0	0	0	0	0	0	0	0	0	0
Z. Cytoskeleton	0	0	0	0	0	0	1	0	0	0
Metabolism										
C. Energy production and conversion	344	223	328	268	504	78	215	214	506	235
E. Amino acid transport and metabolism	392	268	740	508	485	129	421	304	520	364
F. Nucleotide transport and metabolism	93	68	91	86	103	55	88	67	97	72
G. Carbohydrate transport and metabolism	427	213	524	247	416	72	234	101	275	188
H. Coenzyme transport and metabolism	191	131	233	166	205	87	191	156	197	168
I. Lipid transport and metabolism	182	111	269	168	498	91	169	114	415	169

P. Inorganic ion transport and metabolism	179	118	286	239	282	49	205	143	303	186
Q. Secondary metabolites biosynthesis, transport and catabolism	148	70	202	123	403	57	130	63	258	100
Poorly characterized										
R. General function prediction only	833	473	772	462	821	127	544	334	697	418
S. Function unknown	477	308	643	420	347	81	418	278	617	348

SUS: "*Solibacter usitatus*", strain Ellin 6076

345: "*Koreobacter versatilis*", strain Ellin 345

MLOT: *Mesorhizobium loti* MAFF303099

BNC1: *Mesorhizobium* sp. BNC1

MC2 155: *Mycobacterium smegmatis* MC2 155

MLEP: *Mycobacterium leprae* TN

HCHE: *Hahella chejuensis* KCTC 2396

MED92: *Oceanospirillum* sp. MED92

REU: *Ralstonia eutropha* H16

RSO: *Ralstonia solanacearum* UW551

NOTE: Some genes are present in more than one COG category

Table S9. Distribution of genes in COG categories for *Acidobacteria* strains Ellin 6076 and Ellin 345. Only categories with pronounced differences are shown. Shaded categories depict at least a two-fold increase in gene content. The complete list of COG categories and comparisons can be obtained through the Integrated Microbial Genomes system (<http://img.jgi.doe.gov/cgi-bin/pub/main.cgi>).

COG CATEGORY	Strain Ellin 6076	Strain Ellin 345	Fold Increase+
Information storage and processing			
(A) RNA processing and modification	0	1	0
(B) Chromatin structure and dynamics	3	2	1.5
(J) Translation	183	164	1.1
Asp-tRNA ^{Asn} /Glu-tRNA ^{Gln} amidotransferase A subunit and related	13	6	2.2
(K) Transcription	471	295	1.6
Serine/threonine protein kinase	79	30	2.6
DNA-directed RNA polymerase, sigma 24 homolog	70	28	2.5
Transcriptional regulators	114	53	2.2
Response regulators	141	119	1.2
(L) Replication, recombination, repair	328	158	2.1
Serine/threonine protein kinase	79	30	2.6
Site-specific recombinase XerD	32	6	5.3
Transposase and inactivated derivatives	64	10	6.4
Cellular processes			
(D) Cell cycle control, mitosis, meiosis	26	25	1.0
(M) Cell wall/membrane biogenesis	444	257	1.7
Glycosyltransferase	36	15	2.4
4-amino-4-deoxy-L-arabinose transferase and related	26	7	3.7
Glycosyltransferases involved in cell wall biogenesis	36	15	2.4
L-alanine-DL-glutamate epimerase and related	25	4	6.3
Outer membrane protein	40	17	2.4
Nucleoside-diphosphate-sugar epimerases	29	20	1.5
ABC-type transport system, involved in lipoprotein release, permease	17	11	1.6
Membrane-fusion protein	14	8	1.8
Periplasmic protease	13	2	6.5
Endopolygalacturonase	11	0	> 11.0
Dihydropicolinate synthase/N-acetylneuraminate lyase	9	2	4.5
Cell division protein FtsI/penicillin-binding protein 2	7	4	1.8
Outer membrane protein/protective antigen OMA87	6	3	2.0
ABC-type polysaccharide/polyol phosphate export system, permease	6	0	> 6.0
ABC-type polysaccharide/polyol phosphate export system, ATPase	6	0	> 6.0
Periplasmic protein TonB	6	11	NA
Soluble lytic murein transglycolase and related regulatory proteins	6	4	1.5
ADP-heptose:LPS heptosyltransferase	5	2	2.5
Membrane proteins related to metalloendopeptidases	5	1	5.0
Small-conductance mechanosensitive channel	5	2	2.5
Sortase	6	1	6.0
Lipid A core-O-antigen ligase and related	4	2	2.0
(N) Cell motility	101	84	1.2
Tfp pilus assembly protein PilF	26	10	2.6
Type II secretory pathway, pseudopilin PulG	5	9	NA
Type II secretory pathway, component PulD	4	2	2.0
Methyl-accepting chemotaxis protein	4	3	1.3
Flp pilus assembly protein TadC	3	0	> 3.0
Predicted periplasmic or secreted lipoprotein	3	0	> 3.0
(O) Posttranslational modification, protein turnover, chaperones	183	134	1.4
Perirredoxin	21	9	2.3
Protein disulfide isomerase	8	4	2.0
ATPases with chaperone activity, ATP-binding subunit	8	3	2.7
Trypsin-like serine proteases, typically periplasmic, C-terminal PDZ	8	4	2.0
Peptidyl-prolyl cis-trans isomerase (rotamase) – cyclophilin family	7	2	3.5
Molecular chaperone	10	5	2.0
(T) Signal transduction mechanisms	421	316	1.3
Serine/threonine protein kinase	79	30	2.6
Response regulators	141	119	1.2
Signal transduction histidine kinase	88	58	1.5
Antirepressor regulating drug resistance, signal transduction comp.	12	3	4.0
Bacteriophytochrome	9	2	4.5
DnaK suppressor protein	7	2	3.5
(U) Intracellular trafficking, secretion	174	110	1.6

Tfp pilus assembly protein PilF	26	10	2.6
Outer membrane protein	25	10	2.5
Periplasmic component of Tol biopolymer transport system	18	7	2.6
Flp pilus assembly protein TadD	17	9	1.9
Flp pilus assembly protein, ATPase CpaE	4	0	> 4.0
Flp pilus assembly protein, ATPase CpaF	4	0	> 4.0
Type II secretory pathway, component PulD	4	2	2.0
Signal peptidase I	4	3	1.3
Flp pilus assembly protein TadB	3	0	> 3.0
Flp pilus assembly protein CpaB	3	0	> 3.0
Flp pilus assembly protein TadC	3	0	> 3.0
Flp pilus assembly protein, secretin CpaC	2	0	> 2.0
Flp pilus assembly protein TadG	2	0	> 2.0
Flp pilus assembly protein, pilin Flp	2	0	> 2.0
Flp pilus assembly protein, protease CpaA	1	0	> 1.0
(V) Defense mechanisms	236	116	2.0
ABC-type antimicrobial peptide transport system	124	54	2.6
ABC-type multidrug transport system	44	21	2.1
Beta-lactamase class C and other penicillin binding proteins	23	13	1.8
Multidrug resistance efflux pump	18	9	2.0
Cation/multidrug efflux pump	19	8	2.4
Lantibiotic modifying enzyme	2	0	> 2.0
(W) Extracellular structures	0	0	0
(Z) Cytoskeleton	0	0	0
Metabolism			
(C) Energy production, conversion	344	223	1.5
Predicted oxidoreductases	20	8	2.5
Aerobic-type carbon monoxide dehydrogenase, small subunit	9	3	3.0
Aerobic-type carbon monoxide dehydrogenase, large subunit	8	3	2.7
Glycosyltransferases, related to UDP-glucuronosyltransferase	8	3	2.7
FAD/FMN-containing dehydrogenases	8	2	4.0
Carbon dioxide conc. mechanism/carboxysome shell proteins	9	0	> 9.0
FOG:HEAT repeat	6	0	> 6.0
Rieske Fe-S protein	5	0	> 5.0
Predicted acetamidase/formamidase	4	0	> 4.0
Cytochrome b subunit	4	0	> 4.0
Cytochrome c2	3	0	> 3.0
(E) Amino acid transport, metabolism	392	268	1.5
Dipeptidyl aminopeptidases/acylaminoacylpeptidases	25	10	2.5
Acetylornithine deacetylase and related	17	5	3.4
Lysophospholipase L1 and related esterases	16	1	16.0
Threonine dehydrogenase and related Zn-dependent dehydrogenases	12	5	2.4
Aspartate/tyrosine/aromatic aminotransferase	9	6	1.5
Dihydropicolinate synthase/N-acetylneuraminate lyase	9	2	4.5
Choline dehydrogenase and related	8	2	4.0
Xaa-Pro aminopeptidase	7	3	3.5
Histidinol phosphate/aromatic aminotransferase	6	3	2.0
Thiamine pyrophosphate-requiring enzymes	6	3	2.0
Threonine dehydratase	5	2	2.5
Spermidine synthase	5	0	> 5.0
Asparagine synthase (glutamate hydrolyzing)	4	0	> 4.0
ABC-type spermidine/putrescine transport systems, ATPase	3	1	3.0
Predicted ornithine cyclodeaminase	3	0	> 3.0
Methionine synthase II (cobalamin-independent)	3	0	> 3.0
(F) Nucleotide transport, metabolism	93	68	1.4
ADP-ribose pyrophosphatase	6	2	3.0
Inosine-uridine nucleoside N-ribohydrolase	3	0	> 3.0
5'-nucleotidase/2',3'-cyclic phosphodiesterase and related esterases	2	0	> 2.0
(G) Carbohydrate transport, metabolism	427	213	2.0
Sugar phosphate isomerases/epimerases	41	6	6.8
Arabinose efflux permease	27	12	2.3
Glucose dehydrogenase	26	0	> 26.0
Nucleoside-diphosphate-sugar epimerases	29	20	1.5
Sugar phosphate permease	16	6	2.7
Gluconolactonase	13	2	6.5
Sugar kinases, ribokinase family	10	3	3.3
Beta-galactosidase/beta-glucuronidase	9	4	2.3
Alpha-L-arabinofuranosidase	8	2	4.0
Glycosyltransferases, related to UDP-glucuronosyltransferase	8	3	2.7
Alpha-L-fucosidase	7	1	7.0

Glucose/sorbose dehydrogenases	6	1	6.0
ABC-type polysaccharide/polyol phosphate export system, permease	6	0	> 6.0
ABC-type polysaccharide/polyol phosphate export system, ATPase	6	0	> 6.0
Fucose permease	6	3	2.0
Predicted xylanase/chitin deacetylase	6	5	1.2
2,4-dihydroxyhept-2-ene-1,7-dioic acid aldolase	5	0	> 5.0
Beta-glucosidase and related glycosidases	5	8	NA
Mannose-6-phosphate isomerase	5	3	1.7
Ribulose-5-phosphate 4-epimerase and related	5	3	1.7
Beta-xylosidase	4	0	> 4.0
Beta-galactosidase	4	1	4.0
Beta-1,4-xylanase	3	0	> 3.0
Hydroxypyruvate isomerase	3	0	> 3.0
D-mannonate dehydratase	3	0	> 3.0
Ribose/xylose/arabinose/galactoside ABC-type transport, permease	3	0	> 3.0
(H) Coenzyme transport, metabolism	191	131	1.5
Methylase involved in ubiquinone/menaquinone biosynthesis	37	19	2.0
2-polyprenyl-3-methyl-5-hydroxy-6-methoxy-1,4-benzoquinol methylase	18	4	4.5
Mg-chelatase subunit ChlD	8	4	2.0
Demethylmenaquinone methyltransferase	7	0	> 7.0
(I) Lipid transport, metabolism	182	110	1.7
Dehydrogenases with different specificities	39	15	2.6
Esterase/lipase	11	5	2.2
Carboxylesterase type B	9	0	> 9.0
Acyl-CoA synthetases (AMP-forming)	8	4	2.0
(P) Inorganic ion transport, metabolism	179	118	1.5
Arylsulfatase A and related enzymes	18	0	> 18.0
Enterochelin esterase and related enzymes	16	2	8.0
Outer membrane receptor for ferrienterochelin and colicins	7	3	2.3
Mn2+ and Fe2+ transporters of the NRAMP family	6	3	3.0
Cytochrome c peroxidase	6	0	> 6.0
Putative silver efflux pump	3	1	3.0
(Q) Secondary metabolites	148	70	2.1
Dehydrogenases with different specificities	39	15	2.6
Imidazolonepropionase and related amidohydrolases	21	11	1.9
Dienelactone hydrolase and related enzymes	8	1	8.0
Acyl-CoA synthetases (AMP-forming)	8	4	2.0
Carbon dioxide conc. mechanism/carboxysome shell proteins	9	0	> 9.0
N-acyl-D-aspartate/D-glutamate deacylase	5	2	2.5
Acyl carrier protein	5	3	1.7
Protein involved in biosynthesis of mitomycin antibiotics/fumonisin	4	0	> 4.0
Predicted enzyme involved in methoxymalonyl-ACP biodynesis	4	1	4.0
3-oxoacyl-(acyl-carrier-protein) synthase	4	2	2.0
Polyketide synthase modules and related	3	1	3.0
Poorly characterized			
(R) General function prediction	833	472	1.8
(S) Function unknown	477	308	1.5
(*) Not assigned	2943	1612	1.8

+ Not normalized for genome size

Table S10. Functional categories in Ellin 6076

Category	Fold increase in Ellin6076 compared to Ellin345 (# genes)	Possible functions
Greater than 4-fold increase		
Replication, recombination, repair		
Transposase and inactivated derivatives	6.4	Movement of DNA within or between genomes DNA recombination (1)
Site-specific recombinase XerD	5.3	
Total IS elements	(123 in 6076 29 in 345)	
Cell wall/membrane		
4-amino-4-deoxy-L-arabinose transferase and related	3.7 (26 in 6076 7 in 345)	Defense mechanisms: Final step in the polymyxin-resistance pathway (2, 3)
Glycosyl transferase family 39		
L-alanine-DL-glutamate epimerase and related enzymes	6.3 (25 in 6076 4 in 345)	Aromatic acid catabolism (4)
Mandelate racemase/muconate lactonizing enzyme		
Periplasmic protease	6.5 (14 in 6076 2 in 345)	Protein/peptide degradation (5)
peptidase S41		
Endopolygalacturonase	> 11.0 (11 in 6076)	Polygalacturonase (pectinase) catalyses the random hydrolysis of 1,4-alpha-D-galactosiduronic linkages in pectate and other galacturonans (6). Important for bacterial invasion of plants (7).
Glycoside hydrolase family 28		
Dihydropicolinate synthase (DHDPS)/N-acetylneuraminate lyase (NAL) family	4.5 (9 in 6076)	In bacteria, NAL is involved in the regulation of intracellular sialic acid, where it

4.2.1.52 Dihydropicolinate synthase	2 in 345)	catalyzes the cleavage of N-acetylneuraminate to produce pyruvate and N-acetyl-D-mannosamine (8). DHDPS catalyzes the first step in the biosynthesis of lysine via the diaminopimelate pathway (9).
ABC_type polysacch./polyol phosphate export system, permease	>6.0 (6 in 6076)	Exopolysaccharides participate bacterial interactions with plants, protect against dessication, and promote adhesion and biofilm formation (10).
ABC-2 type transporter		
ABC_type polysacch./polyol phosphate export system, ATPase	>6.0 (6 in 6076)	
Membrane proteins related to metalloendopeptidases	5.0	Proteolysis
peptidase M23B	(5 in 6076 1 in 345)	
Sortase	6.0 (6 in 6076 1 in 345) 4 genes in 6076 are not found in 345: "LPXTG-site transpeptidase (sortase) family protein". The rest are "GCN5-related N-acetyltransferase"	Sortase is a surface protein transpeptidase that acts on surface proteins and anchors them to the cell wall (Mazmanian et al, 1999).
Signal Transduction		
Antirepressor regulating drug resistance, signal transduction comp.	4.0 (12 in 6076)	Defense mechanisms: Drug resistance (Brakstad and Maeland, 1997)
Peptidase M56, BlaR1	3 in 345)	
Bacteriophytochrome	4.5 (9 in 6076)	Bacteriophytochromes are phytochrome-like proteins that mediate photosensory responses in various bacteria according to their light environment (Jaubert et al, 2007).
multi-sensor signal transduction histidine kinase	2 in 345)	

Intracellular trafficking, secretion		
Flp pilus assembly protein ATPase CpaE	>4.0	Not obvious why having extras of these would be beneficial
Flp pilus assembly protein ATPase CpaF	>4.0	See above
Energy production, conversion		
FAD/FMN-containing dehydrogenases	4.0 (8 in 6076)	General metabolism
FAD linked oxidase	2 in 345)	
Carbon dioxide conc. mechanism/carboxysome shell proteins	> 9.0	Some non-autotrophic organisms form polyhedral organelles resembling carboxysomes found in autotrophs. Polyhedral organelles in non-autotrophic organisms are involved in coenzyme B12-dependent 1,2-propanediol utilization and ethanolamine utilization (11, 12). Ellin 6076 did have homologs of <i>S. enterica</i> PduJ (propanediol utilization protein) and PduP (putative CoA-dependent propionaldehyde dehydrogenase).
3 annotated as Ethanolamine utilization protein EutN/carboxysome structural protein CcmI		FOG:HEAT repeat proteins include the lyase enzymes that specifically attach particular phycobilins to apophycobiliprotein subunits (13, 14).
6 annotated as microcompartments protein		Rieske iron-sulfur proteins are subunits of cytochrome bc-type complexes or oxygenases, and may allow an organism's electron transfer chains to adapt to changing environmental conditions (15).
FOG:HEAT repeat protein	> 6.0	Amide metabolism (16).
PBS lyase HEAT domain protein repeat-containing protein	6 in 6076	Part of the cytochrome bc1 complex, composed of membrane proteins that catalyze the
Rieske Fe-S protein	> 5.0	
Rieske (2Fe-2S) domain protein	5 in 6076	
Predicted acetamidase/formamidase	>4.0	
Cytochrome b subunit	>4.0	

Cytochrome b/b6

oxidation of ubihydroquinone and the reduction of cytochrome c in bacterial photosynthetic and respiratory chains (17).

Amino acid transport, metabolism

Lysophospholipase L1 and related esterases	16.0 (16 in 6076)	Lipid metabolism
lipolytic enzyme, G-D-S-L family	1 in 345)	
Dihydropicolinate synthase/N-acetylneuraminate lyase	4.5	See above
Annotated as dihydrodipicolinate synthetase	(9 in 6076 2 in 345)	
Choline dehydrogenase and related	4.0	Metabolism (18)
Most annotated as glucose-methanol-choline oxidoreductase	(8 in 6076 2 in 345)	
Spermidine synthase	> 5.0	Involved in the fifth (last) step in the biosynthesis of spermidine from arginine and methionine (19).
Asparagine synthase (glutamine-hydrolyzing) (EC:6.3.5.4)	> 4.0	The enzyme asparagine synthetase produces asparagine, AMP, glutamate, and pyrophosphate from aspartate, glutamine, and ATP (20).

Carbohydrate transport, metabolism

Sugar phosphate isomerases/epimerases	6.8	This TIM alpha/beta barrel structure is found in xylose isomerase, endonuclease IV, and bacterial myo-inositol catabolism proteins (21, 22).
Xylose isomerase domain protein TIM barrel	(41 in 6076 6 in 345)	
Glucose dehydrogenase	> 26.0	Quinoprotein glucose dehydrogenase (pyrrolo-quinoline quinone) functions in glucose and glucose-1-phosphate degradation (23-25).
Quinoprotein glucose dehydrogenase (EC:1.1.5.2)	26 in 6076 none in 345	

Gluconolactonase (EC 3.1.1.17)	6.5 (13 in 6076 2 in 345)	Acts in the pathways: 6-phosphogluconate synthesis via gluconate Nonphosphorylated Entner-Doudoroff pathway Semiphosphorylated Entner-Doudoroff pathway (26).
Alpha-L-fucosidase glycoside hydrolase, family 29 (alpha-L- fucosidase) (EC:3.2.1.51)	7.0 (7 in 6076 1 in 345)	Hydrolyzes the alpha-1,6-linked fucose at the reducing-end N-acetylglucosamine in the carbohydrate parts of glycoproteins (27).
Glucose/sorbose dehydrogenases	6.0 (6 in 6076 1 in 345)	Oxidizes glucose to gluconolactone and uses PQQ as a cofactor (24, 25).
ABC_type polysacch./polyol phosphate export system, permease	>6.0	See above
ABC-type polysaccharide/polyol phosphate export system, permease and ATPase	> 6.0	See above
2,4-dihydroxyhept-2-ene-1,7-dioic acid aldolase HpcH/HpaI aldolase (EC:4.1.2.20)	> 5.0	Part of homoprotocatechuate pathway (28).
glycoside hydrolase, family 43 (endo-1,5- alpha-L-arabinosidase)	6.0 (6 in 6076 1 in 345)	Arabinose degradation Xylan degradation (29)
glycoside hydrolase family 39 (beta- xylosidase)	4 genes in 6076 encode endo- 1,5-alpha-L-arabinosidase, the rest encode beta-xylosidases	
Beta-galactosidase (3.2.1.23)	4.0	Hydrolysis of terminal non-reducing beta-D- galactose residues in beta-D-galactosides (29)
Coenzyme transport, metabolism		
2-polyprenyl-3-methyl-5-hydroxy-6-methoxy- 1,4-benzoquinol methylase	4.5 (20 in 6076 4 in 345)	UbiG Last step in CoQ (ubiquinone) biosynthesis. Ubiquinone functions in the respiratory electron transport chain of prokaryotes (30).
annotated as Methyltransferase type 12		

Methyltransferase type 11		
Demethylmenaquinone methyltransferase	> 7.0 7 in 6076	Demethylmenaquinone methyltransferases convert dimethylmenaquinone to menaquinone in the final step of menaquinone biosynthesis. Menaquinone functions in the electron transport chain (31).
Lipid transport, metabolism		
Carboxylesterase type B	> 9.0 9 in 6076	Ester hydrolysis (32)
Inorganic ion transport, metabolism		
Arylsulfatase A and related enzymes	> 18.0 18 in 6076	Sulfur metabolism. Sulphatases are enzymes that hydrolyze various sulphate esters (33). Ester hydrolysis (32)
Enterochelin esterase and related enzymes	8.0 (16 in 6076 2 in 345)	
Cytochrome c peroxidase	> 6.0 6 in 6076	Bacterial cytochrome c peroxidases reduce hydrogen peroxide without the need to generate semi-stable free radicals (34).
Secondary metabolites		
Dienelactone hydrolase and related enzymes	8.0	Dienelactone hydrolases play a crucial role in chlorocatechol degradation via the modified ortho cleavage pathway (35).
Carboxymethylenebutenolidase (EC:3.1.1.45)	(8 in 6076	
dienelactone hydrolase (EC:3.1.1.45)	1 in 345)	See above
Carboxysome shell proteins	See above	
Protein involved in biosynthesis of mitomycin antibiotics/fumonisin	>4.0 4 in 6076	This family is made up of several eukaryotic phytanoyl-CoA dioxygenase (PhyH) proteins, ectoine hydroxylases and a number of bacterial deoxygenases (36).
Phytanoyl-CoA dioxygenase		
Predicted enzyme involved in methoxymalonyl-ACP biosynthesis	4.0 (4 in 6076	Secondary metabolism – antibiotic biosynthesis

FkbH like protein	1 in 345)	
Less than 4-fold increase		
Regulation		
Sigma-24 (sigmaE) ECF (74 genes in 6076)	2.6	Adaptation to environmental conditions (37); Survival under nutrient deprivation and resistance to oxidative stress (38); Carotenoid biosynthesis (39); Cell envelope stress tolerance and response (40-43); Response to deleterious environmental conditions, ie, from glucose to secondary carbon sources (44)
Anti-sigma factors (14 in 6076)	1.75	
Two-component systems		Cell envelope stress tolerance and response (40)
In 6076: 109 histidine kinase 68 response regulator receiver		
in 345: 70 histidine kinase 63 response regulator		
Serine/threonine protein kinases (some are potential anti-sigma factors)		Serine/threonine protein kinases (EC 2.7.11.1) phosphorylate the OH group of serine or threonine (which have similar sidechains), and play a role in various cellular processes (45).
35 in 345 82 in 6076		

TPR repeat proteins

122 in 6076

75 in 345

The tetratricopeptide repeat (TPR) is found in a wide range of proteins, and mediates protein-protein interactions and the assembly of multiprotein complexes (46, 47).

References:

1. Sirois S & Szatmari G (1995) Detection of XerC and XerD recombinases in gram-negative bacteria of the family Enterobacteriaceae. *J. Bacteriol.* 177:4183-4186.
2. Bretscher LE, Morrell MT, Funk AL, & Klug CS (2006) Purification and characterization of the L-Ara4N transferase protein ArnT from *Salmonella typhimurium*. *Protein Expr. Purif.* 46:33-39.
3. Trent MS, Ribeiro AA, Lin S, Cotter RJ, & Raetz CRH (2001) An inner membrane enzyme in *Salmonella* and *Escherichia coli* that transfers 4-amino-4-deoxy-L-arabinose to lipid A: induction on polymyxin-resistant mutants and role of a novel lipid-linked donor. *J. Biol. Chem.* 276:43122-43131.
4. Neidhart DJ, Kenyon GL, Gerlt JA, & Petsko GA (1990) Mandelate racemase and muconate lactonizing enzyme are mechanistically distinct and structurally homologous. *Nature* 347:692-694.
5. Rawlings ND & Barrett AJ (1994) Families of serine peptidases. *Meth. Enzymol.* 244:19-61.
6. Markovic O & Janecek S (2001) Pectin degrading glycoside hydrolases of family 28: sequence-structural features, specificities and evolution. *Protein Eng.* 14:615-631.
7. Ljunggren I & Fåhræus G (1961) The Role of Polygalacturonase in Root-Hair Invasion by Nodule Bacteria *J. Gen. Microbiol.* 26:521-528.
8. Schauer R (1982) Chemistry, metabolism, and biological functions of sialic acids. *Adv. Carbohydr. Chem. Biochem.* 40:131-234.
9. Gilvarg C (1960) Biosynthesis of diaminopimelic acid. *Fin. Lakaresallsk Handl.* 19:948-952.
10. Weiner R, Langille S, & Quintero E (1995) Structure, function and immunochemistry of bacterial exopolysaccharides. *J. Ind. Microbiol.* 15:339-346.
11. Havemann GD, Sampson EM, & Bobik TA (2002) PduA is a shell protein of polyhedral organelles involved in coenzyme B(12)-dependent degradation of 1,2-propanediol in *Salmonella enterica* serovar typhimurium LT2. *J. Bacteriol.* 184:1253-1261.

12. Kofoed E, Rappleye C, Stojiljkovic I, & Roth J (1999) The 17-gene ethanolamine (eut) operon of *Salmonella typhimurium* encodes five homologues of carboxysome shell proteins. *J. Bacteriol.* 181:5317-5329.
13. Dolganov N & Grossman AR (1999) A polypeptide with similarity to phycocyanin alpha-subunit phycocyanobilin lyase involved in degradation of phycobilisomes. *J. Bacteriol.* 181:610-617.
14. Fairchild CD & Glazer AN (1994) Oligomeric structure, enzyme kinetics, and substrate specificity of the phycocyanin alpha subunit phycocyanobilin lyase. *J. Biol. Chem.* 269:8686-8694.
15. Schneider D & Schmidt CL (2005) Multiple Rieske proteins in prokaryotes: where and why? *Biochim. Biophys. Acta.* 1710:1-12.
16. Mills J, Greenwood JA, & Jones CW (2006) The effect of growth in continuous culture at various input C:N ratios on the expression of proteins involved in the transport and metabolism of methanol and short-chain amides by *Methylophilus methylotrophus*. *FEMS Microbiol. Lett.* 161:173-178.
17. Crofts AR (2004) The cytochrome bc1 complex: function in the context of structure. *Annu. Rev. Physiol.* 66:689-733.
18. Cavener DR (1992) GMC oxidoreductases. A newly defined family of homologous proteins with diverse catalytic activities. *J. Mol. Biol.* 223:811-814.
19. Wu H, *et al.* (2007) Structure and mechanism of spermidine synthases. *Biochemistry* 46:8331-8339.
20. Yoshida K, Fujita Y, & Ehrlich SD (1999) Three asparagine synthetase genes of *Bacillus subtilis* *J. Bacteriol.* 181:6081-6091.
21. Fenn TD, Ringe D, & Petsko GA (2004) Xylose isomerase in substrate and inhibitor michaelis states: atomic resolution studies of a metal-mediated hydride shift. *Biochemistry* 43:6464-6474.
22. Fry J, Wood M, & Poole PS (2001) Investigation of myo-inositol catabolism in *Rhizobium leguminosarum* bv. viciae and its effect on nodulation competitiveness. *Mol. Plant Microbe Interact.* 14:1016-1025.
23. Duine JA & Jongejan JA (1989) Quinoproteins, enzymes with pyrrolo-quinoline quinone as cofactor. *Annu. Rev. Biochem.* 58:403-426.
24. Gallop PM, Paz MA, Fluckiger R, & Kagan HM (1989) PQQ, the elusive coenzyme. *Trends Biochem. Sci.* 14:343-346.
25. Oubrie A, *et al.* (1999) Structure and mechanism of soluble quinoprotein glucose dehydrogenase. *EMBO J.* 18:5187-5194.
26. Kanagasundaram V & Scopes R (1992) Isolation and characterization of the gene encoding gluconolactonase from *Zymomonas mobilis*. *Biochim. Biophys. Acta.* 1171:198-200.
27. Fisher KJ & Aronson Jr. NN (1989) Isolation and sequence analysis of a cDNA encoding rat liver alpha-L-fucosidase. *Biochem. J.* 264:695-701.

28. Stringfellow JM, Turpin B, & Cooper RA (1995) Sequence of the *Escherichia coli* C homoprotocatechuic acid degradative operon completed with that of the 2,4-dihydroxyhept-2-ene-1,7-dioic acid aldolase-encoding gene (hpcH). *Gene* 166:73-76.
29. Davies G & Henrissat B (1995) Structures and mechanisms of glycosyl hydrolases. *Structure* 3:853-859.
30. Meganathan R (2001) Ubiquinone biosynthesis in microorganisms. *FEMS Microbiol. Lett.* 203:131-139.
31. Koike-Takeshita A, Koyama T, & Ogura K (1997) Identification of a novel gene cluster participating in menaquinone (vitamin K₂) biosynthesis. Cloning and sequence determination of the 2-heptaprenyl-1,4-naphthoquinone methyltransferase gene of *Bacillus stearothermophilus*. *J. Biol. Chem.* 272:12380-12383.
32. Prim N, Pastor FI, & Diaz P (2001) Cloning and characterization of a bacterial cell-bound type B carboxylesterase from *Bacillus* sp. BP-7. *Curr. Microbiol.* 42:237-240.
33. Boltes I, *et al.* (2001) 1.3 Å structure of arylsulfatase from *Pseudomonas aeruginosa* establishes the catalytic mechanism of sulfate ester cleavage in the sulfatase family. *Structure* 9:483-489.
34. Fulop V, Ridout CJ, Greenwood C, & Hajdu J (1995) Crystal structure of the di-haem cytochrome c peroxidase from *Pseudomonas aeruginosa*. *Structure* 3:1225-1233.
35. Schlömann M, Ngai KL, Ornston LN, & Knackmuss HJ (1993) Dienelactone hydrolase from *Pseudomonas cepacia*. *J. Bacteriol.* 175:2994-3001.
36. Prabhu J, Schauwecker F, Grammel N, Keller U, & Bernhard M (2004) Functional expression of the ectoine hydroxylase gene (thpD) from *Streptomyces chrysomallus* in *Halomonas elongata*. *Appl. Environ. Microbiol.* 70:3130-3132.
37. Helmann JD (2002) The extracytoplasmic function (ECF) sigma factors. *Adv Microb Physiol.* 46:47-100.
38. Testerman TL, *et al.* (2002) The alternative sigma factor sigmaE controls antioxidant defences required for *Salmonella* virulence and stationary-phase survival. *Mol. Microbiol.* 43:771-782.
39. Thirunavukkarasu N, *et al.* (2008) An extra-cytoplasmic function sigma factor and anti-sigma factor control carotenoid biosynthesis in *Azospirillum brasilense*. *Microbiology* 154:2096-2105.
40. Biswas I, Drake L, Erkina D, & Biswas S (2008) Involvement of sensor kinases in the stress tolerance response of *Streptococcus mutans*. *J. Bacteriol.* 190:68-77.
41. Lemos JA & Burne RA (2008) A model of efficiency: stress tolerance by *Streptococcus mutans*. *Microbiology* 154:3247-3255.
42. Suntharalingam P, Senadheera MD, Mair RW, Lévesque CM, & Cvitkovitch DG (2009) The LiaFSR system regulates the cell envelope stress response in *Streptococcus mutans*. *J. Bacteriol.* 191:2973-2984.

43. Wecke T, Veith B, Ehrenreich A, & Mascher T (2006) Cell envelope stress response in *Bacillus licheniformis*: integrating comparative genomics, transcriptional profiling, and regulon mining to decipher a complex regulatory network. *J. Bacteriol.* 188:7500-7511.
44. Kenyon WJ, Thomas SM, Johnson E, Pallen MJ, & Spector MP (2005) Shifts from glucose to certain secondary carbon-sources result in activation of the extracytoplasmic function sigma factor sigmaE in *Salmonella enterica* serovar Typhimurium. *Microbiology* 151:2373-2383.
45. Manning G, Plowman GD, Hunter T, & Sudarsanam S (2002) Evolution of protein kinase signaling from yeast to man. . *Trends Biochem. Sci.* 27:514-520.
46. Blatch GL & Lässle M (1999) The tetratricopeptide repeat: a structural motif mediating protein-protein interactions. *Bioessays* 21:932-939.
47. Smith RL & Johnson AD (2000) A sequence resembling a peroxisomal targeting sequence directs the interaction between the tetratricopeptide repeats of Ssn6 and the homeodomain of alpha 2. *Proc. Natl. Acad. Sci. USA.* 97:3901-3906.



Table S11. Metabolic enzymes

Category	# genes in Ellin6076	# genes in Ellin345
Oxidoreductase	128	64
Transferase	427	267
Hydrolase	198	123
Lyase	22	16
Isomerase	89	39
Ligase	39	32

Table S12. Central and peripheral aromatic pathways from other large genomes (*Burkholderia xenovorans* LB400 and *Rhodococcus* sp. RHA1)

Pathway	Ellin 6076	Ellin 345
Protocatechuate conversion to TCA cycle intermediates*	yes	no
Vanillate conversion to protocatechuate	no	no
3-hydroxyphenylacetate conversion to protocatechuate	no	no
Catechol conversion to beta-ketoadipate enol-lactone	no	no
Benzoate conversion to catechol	no	no
Benzoate conversion to Benzoyl-CoA	no	no
Phthalate catabolism	no	no
Terephthalate catabolism	no	no
L-phenylalanine degradation via homogentisate	yes	yes
L-tyrosine degradation via homogentisate	yes	yes
Phenol conversion to catechol	no	no

*Genes involved in procatechuate catabolism are homologs of the *Rhodococcus* RHA1 genes. Gene order and orientation in Ellin 6076 were different from RHA1, but some of the flanking genes were also in common. The 6076 gene cluster was flanked by tRNA-Met genes on either side (Acid_R0003 and Acid_R0004).