Phylogenetic and Phenotypic Characterization of the Energy-taxis Receptor Aer in *Pseudomonas*

Sean C. Booth[a]#* and Raymond J. Turner[a]#

Department of Biological Sciences, University of Calgary, Calgary, AB, Canada[a]

#Address correspondence Raymond J. Turner, turnerr@ucalgary.ca and Sean C. Booth, scbooth@ucalgary.ca

* Current address: Singapore Centre for Environmental Life Sciences Engineering, Nanyang Technological University, Singapore

+1-403-220-3581

Running Title: Pseudomonas Energy-Taxis Receptors

Keywords: chemotaxis, *Pseudomonas*, aerotaxis, energy-taxis, phylogenetics, motility

Abstract

Chemotaxis allows bacteria to sense gradients in their environment and respond by directing their swimming. Aer is a receptor that, instead of responding to a specific chemoattractant, allows bacteria to sense cellular energy production and move towards favourable environments. In *Pseudomonas*, the number of apparent Aer homologs differed between the only two species it had been characterized in, *P. aeruginosa* and *P. putida*. Here we combined bioinformatic approaches with deletional mutagenesis in *P. pseudoalcaligenes* KF707 to further characterize Aer. It was determined that the number of Aer homologs varies between 0-4 throughout the *Pseudomonas* genus, and they were phylogenetically classified into 5 subgroups. We also used sequence analysis to show that these homologous receptors differ not in the ligand binding or signal transduction domains, but in the region in between. The most prevalent homolog was thus differentiated from *Escherichia coli* Aer by its different domain architecture. Genetic analysis also indicated that some Aer homologs have likely been subject to horizontal transfer. *P. pseudoalcaligenes* KF707 was unique among species for having three Aer homologs as well as the distinct chemoreceptor Aer-2. Phenotypic characterization in this species showed the most prevalent homolog was key, but not essential for energy-taxis. This study demonstrates that energy-taxis in *Pseudomonas* varies between species and provides a new naming convention and associated phylogenetic details for Aer chemoreceptors.

Importance

Energy-taxis enables *Pseudomonas* to swim towards favourable environments through sensing its cellular energy state via the receptor ‘Aer’. In *Escherichia coli* there is only one version of this gene but *Pseudomonads* appeared to have multiple. Here we show that there are 5

different homologs in *Pseudomonas* and that individual species can have between zero and four Aer homologs. These homologs do not differ in their ligand-binding region or signal transduction module, but in the region between them. Only one homolog has a HAMP domain like *E. coli* Aer, making the other homologs interesting among chemoreceptors for not having this domain. In *P. pseudoalcaligenes* KF707, which has a unique complement of Aer homologs and other chemoreceptors, all Aer homologs influenced energy-taxis.

Introduction

Chemotaxis is the ability to sense and swim along chemical gradients, and is a widespread, important behaviour in bacteria (1). Canonically, it functions by extracellular compounds interacting with membrane-bound chemoreceptors, generally called methyl-accepting chemotaxis proteins (MCPs), causing a phosphorylation signal cascade through CheA and CheY to alter the direction of flagellar rotation (1), which in turn allows the cell to direct swimming through concentration gradients. The first chemoreceptors were characterized in *E. coli*, but now many more have been described, particularly in *Pseudomonas* (2). Many receptors for specific ligands are being identified through a high-throughput approach (3), but some specialized receptors do not recognize extracellular signals.

Energy-taxis is a behavior in bacteria that enables swimming towards optimal environments for producing energy by sensing intracellular signals (4). This can mean regions with higher concentrations of metabolizable carbon sources, or regions with higher oxygen concentrations. This has resulted in the conflation of directed swimming towards oxygen, called aerotaxis, with energy-taxis. This is most evident in the naming of the energy-taxis receptor ‘Aer’. First described in *E. coli* (5, 6), this receptor does not sense oxygen concentrations through direct

binding, instead it detects the cellular energy state through sensing the redox state of a flavin adenine dinucleotide (FAD) co-factor (7). An orthologous receptor was also discovered in *P. aeruginosa*, along with a second distinct protein, deemed Aer-2 (formerly known as McpB), and both contributed to aerotaxis (8). In *P. putida* three *aer*-like genes were found, but only one was important for energy-taxis (9). Unfortunately, this chemoreceptor was given the numerical name 'Aer2' despite it referring to a completely different protein than the previously named Aer-2. In this study, we sought to understand the variation in Aer chemoreceptors in the genus *Pseudomonas* and provide phylogenetic data to rename the multiple Aer homologs. Additionally, we aimed to investigate the functions of Aer and Aer-2 through deletional mutagenesis in a species they have yet to be investigated in, *P. pseudoalcaligenes* KF707.

In *Pseudomonas*, the energy-taxis receptor Aer has been characterized using gene inactivated mutants of 2 species, first in *P. putida* PRS2000 (10), then in *P. aeruginosa* PA01 (8), *P. putida* KT2440 (9) and *P. putida* F1 (11). In *P. putida* KT2440, three genes with high sequence identity to Aer from *P. aeruginosa* PA01 were individually disrupted, and 'Aer2' was the only one found to mediate energy-taxis (9). Though this species had two more potential Aer homologs than *P. aeruginosa*, no comparison of the genes or gene products was made. A better understanding of energy-taxis in *Pseudomonas* and why some species have multiple similar copies of the same chemoreceptor could thus be obtained by studying how the number and amino acid sequences of Aer homologs varies in the genus and which are functional in other species.

In *P. aeruginosa*, Aer-2 was originally implicated with Aer as an aerotaxis receptor (8), but its function remains uncertain, despite its ability to directly bind oxygen (12). The *aer-2* gene is part of the *che2* gene cluster, which is preceded by another chemoreceptor, *mcpA* (13). This gene was demonstrated to mediate positive chemotaxis towards tetrachloroethylene (TCE) and

renamed *cttP* (14). This effect was detectable only in a strain with disruptions to all 3 amino acids receptors (*pctABC*). Compared to other MCPs, CttP has a strange domain architecture; it has no clear ligand binding region and has a C-terminal extension. In *E. coli,* its MCPs can mediate attractant and repellent responses to phenol that do not involve their ligand binding regions (15). This implies the observed taxis to TCE may also have been fortuitous and that CttP has some other function. As *cttP* is located beside the *che2* gene cluster which contains Aer-2, we hypothesized that in addition to Aer, CttP and Aer-2 may also have a role in energy-taxis.

In this study we present a combination of a bioinformatic characterization of Aer throughout the *Pseudomonas* genus with a genetic knockout characterization of three Aer homologs, Aer-2 and CttP in *P. pseudoalcaligenes* KF707. A phylogeny of Aer was built using sequences obtained from 65 *Pseudomonas* species providing insight into the distribution throughout the genus and enabling the definition of 5 'Aer' subgroups. Only *P. pseudoalcaligenes* KF707 had the unique feature of possessing three Aer homologs, Aer-2 and CttP making it the ideal candidate to investigate if these receptors had related functions. Using single and combinatory deletion mutants, all five of these genes were found to have some influence on energy-taxis, with the most common Aer homolog in the genus playing the most important role. Together these results provide a definition of the Aer energy-taxis receptor as a family with varied distribution in *Pseudomonas* and imply some role for Aer-2 and CttP in energy-taxis.

Results

Bioinformatics Results

Previous studies suggested that the number of Aer-like receptors varied between *Pseudomonas* species, but it was unknown how much this number varied, how much their amino

106 acid sequences varied, and how these Aer-like receptors were related phylogenetically. To
107 address these questions and more, the Aer sequences from 65 *Pseudomonas* species were
108 selected for analysis. Species with completely sequenced genomes were selected, but for those
109 with many strains only a few representatives that have been highly studied were included (e.g. *P.*
110 *aeruginosa*). Species with incomplete (draft) sequences were also included in an attempt to
111 ensure representation from all major *Pseudomonas* clades, based on the phylogenies of Bodilis *et*
112 *al* and Gomila *et al* (16, 17). 144 protein sequences were obtained from the NCBI database using
113 the *P. aeruginosa* PA01 Aer sequence (NP_250252.1) as a BLAST (18) query sequence. All hits
114 with >95% sequence coverage were included (E values $<10^{-100}$). These sequences were aligned
115 using COBALT (19) as this alignment algorithm ensures that conserved domains are aligned
116 despite a lack of similarity elsewhere in the sequence. This was ideal for the Aer sequences as,
117 like other chemoreceptors, they are conserved in the C-terminal signaling region but less
118 conserved in the rest of the protein, making the C-terminal region important for the alignment. A
119 maximum likelihood phylogeny was constructed using PHYML (20), as implemented by the
120 South of France Bioinformatics Platform (21). Through this approach, two phylogenies were
121 generated, one using only the *Pseudomonas* sequences and one which included Aer from *E. coli*
122 K12 as an outgroup. These two unrooted phylogenies were examined (Supplementary Figures 1
123 and 2) and the placement of *E. coli* Aer was used to root the *Pseudomonas*-only tree
124 (Supplementary Figure 3). The rooted phylogeny allowed clear delineation of Aer sequences
125 into subgroups.

126 Phylogenetic Grouping of Aer

127 The alignment of *Pseudomonas* Aer sequences was ordered based on the maximum
128 likelihood phylogeny (see alignment and tree in supplementary material). Inspection of this

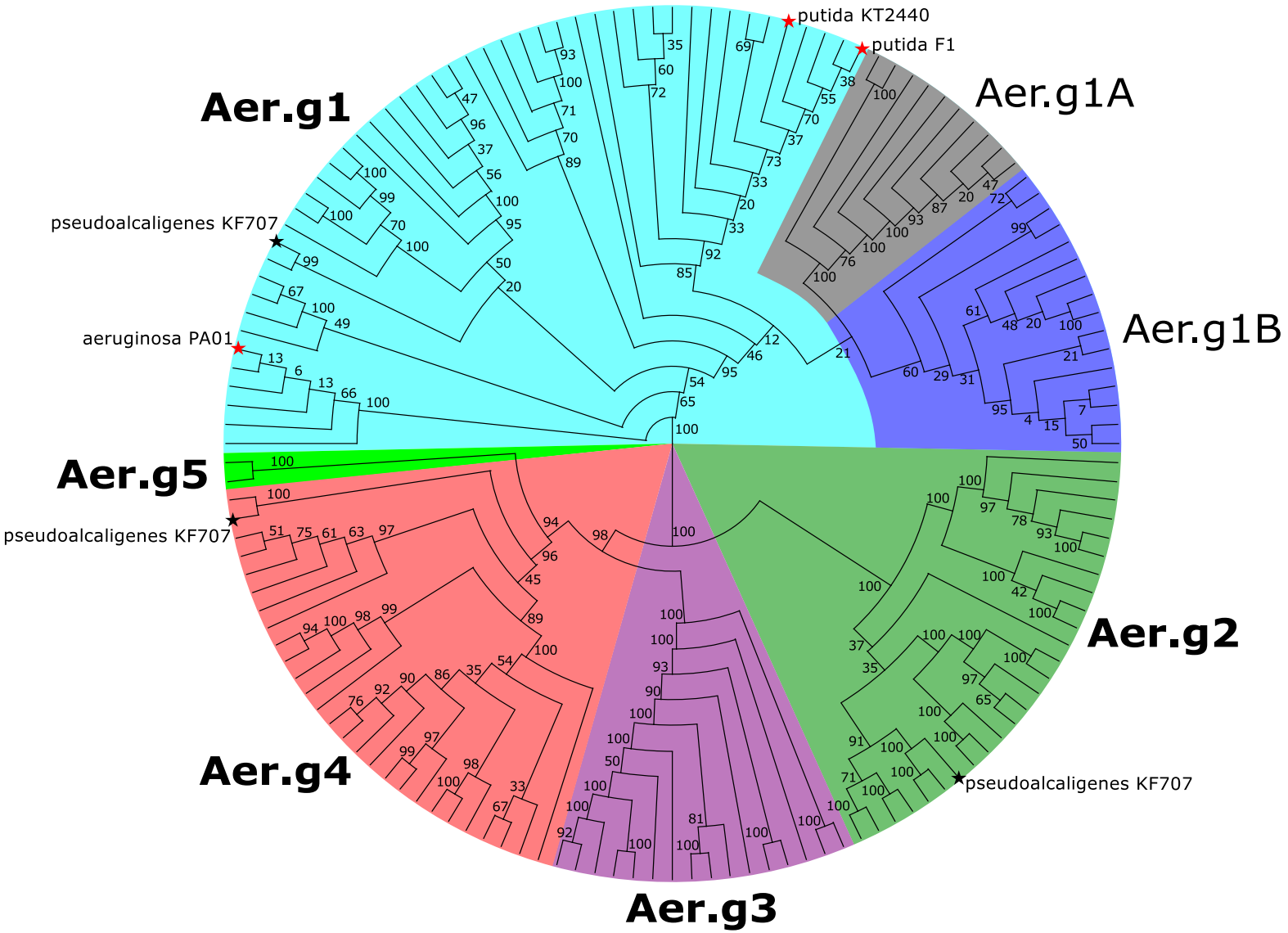


Figure 1: Maximum-Likelihood consensus cladogram showing phylogenetic relationship between Aer protein sequences from *Pseudomonas* species. Sequences were grouped according to branching pattern and inspection of the alignment, and were confirmed in subsequent analysis, see following figures and text for details. Only previously characterized Aer sequences (red stars) and sequences from *P. pseudoalcaligenes* KF707 (black stars) are labeled. For complete tree with branch lengths and all labels see Supplementary Figure 3. Tree was generated unrooted, then the root placed based on comparison between the unrooted tree and a tree rooted to Aer from E. coli, see Supplementary Figures 1 and 2 for details. Numbers along branches indicate bootstrap support values from 100 replicates.

organized alignment implied that there are several sub-families of Aer homologs (Figure 1). Initially, seven groups were identified based on the alignment and tree branching. These groups were sub-aligned by themselves, then analyzed to determine if the group assignments were accurate. Manual inspection and Sequence Harmony / Multi-Relief (SHMR) demonstrated that the seven groups could be reduced to just five clear subgroupings. SHMR is a pair of algorithms that takes as input a pair of alignments and determines, for each AA position, whether that AA is conserved within each group, and whether it is divergent between the two groups, giving each AA position a score (22). A score of 1 indicates perfect within-group conservation and between-group divergence. Each pair of groups was compared in this fashion, then their results compared by determining what percent of AAs were conserved within a group but divergent between the groups (Supplementary Figure 4). AA positions with a multi-relief weight score >0.8 were accepted as fitting these criteria (cutoff is based on the recommendations of the SHMR authors (22)). Most intergroup comparisons indicated that 25-35% of AAs were above the distinction cutoff, indicating the group distinctions were correct. Conversely, two proposed subgroupings of Aer.g1 were demonstrated to differ from the rest of Aer.g1 by much less than any other between-group comparisons as only 10% of AAs were above the cutoff. These two groups were thus included as part of Aer.g1, reducing the total number of groups to five. Throughout this process, Aer.g5 was consistently excluded as it only contains 2 sequences, which are highly divergent from all others.

Distribution of Groups within *Pseudomonas*

To determine the prevalence of each Aer homolog group within the *Pseudomonas* genus, the number of homologs from each group that each species possessed was counted and a hierarchically clustered heatmap was generated (Supplementary Figure 5). Aer-2 and CttP were

also counted and included. This analysis showed that all included species, except *P. denitrificans* ATCC13867, had an Aer.g1 homolog. Conversely, only species related to *P. aeruginosa* had Aer-2 and CttP, which were always found together. There were very few species that had Aer-2/CttP and multiple Aer homologs, only *Pseudomonas* sp 21, *P. nitroreducens* HBP1, *P. resinovorans* NBRC 106553 and *P. pseudoalcaligenes* KF707 fit this category. Duplication of subgroup genes within a single strain were rarer than possession of homologs from multiple groups and only Aer.g1, Aer.g2 and Aer.g3 were duplicated. Aer.g1 was only duplicated in *P. fluorescens* and related subgroups 1-8 (17), and these duplications matched the prior subdivision of Aer.g1 into Aer.g1A and Aer.g1B. Duplications of Aer.g2 occurred only (and always) in *P. stuzeri* and the closely related *P. balearica*. Unlike the other duplications, these were tandem duplications. Duplications of Aer.g3 only occurred in *P. protegens* CHA0 and Pf-5. No single species had a representative from all groups, though *P. mendocina* NK-01 was only missing Aer.g2 and Aer-2/CttP, and *P. pseudoalcaligenes* KF707 was only missing Aer.g3 and Aer.g5.

At the strain level, there were instances of intraspecies differences. Only half of the *P. parafulva* strains had an Aer.g2 homolog and only half of the *P. fluorescens* strains had identical Aer homolog profiles. Three of four *P. putida* strains had the same Aer homologs, but *P. putida* HB3267 had one extra. *P. mendocina* ymp and NK-01 differed as the latter possesses the rare Aer.g5. The other strain with this homolog, *P. pseudoalcaligenes* AD6, was quite different than KF707 as it also did not have an Aer.g2 homolog nor Aer-2/CttP. This was the species with the most inter-strain variation as the third strain included in the analysis, *P. pseudoalcaligenes* CECT5344, only possessed a single Aer.g1 homolog.

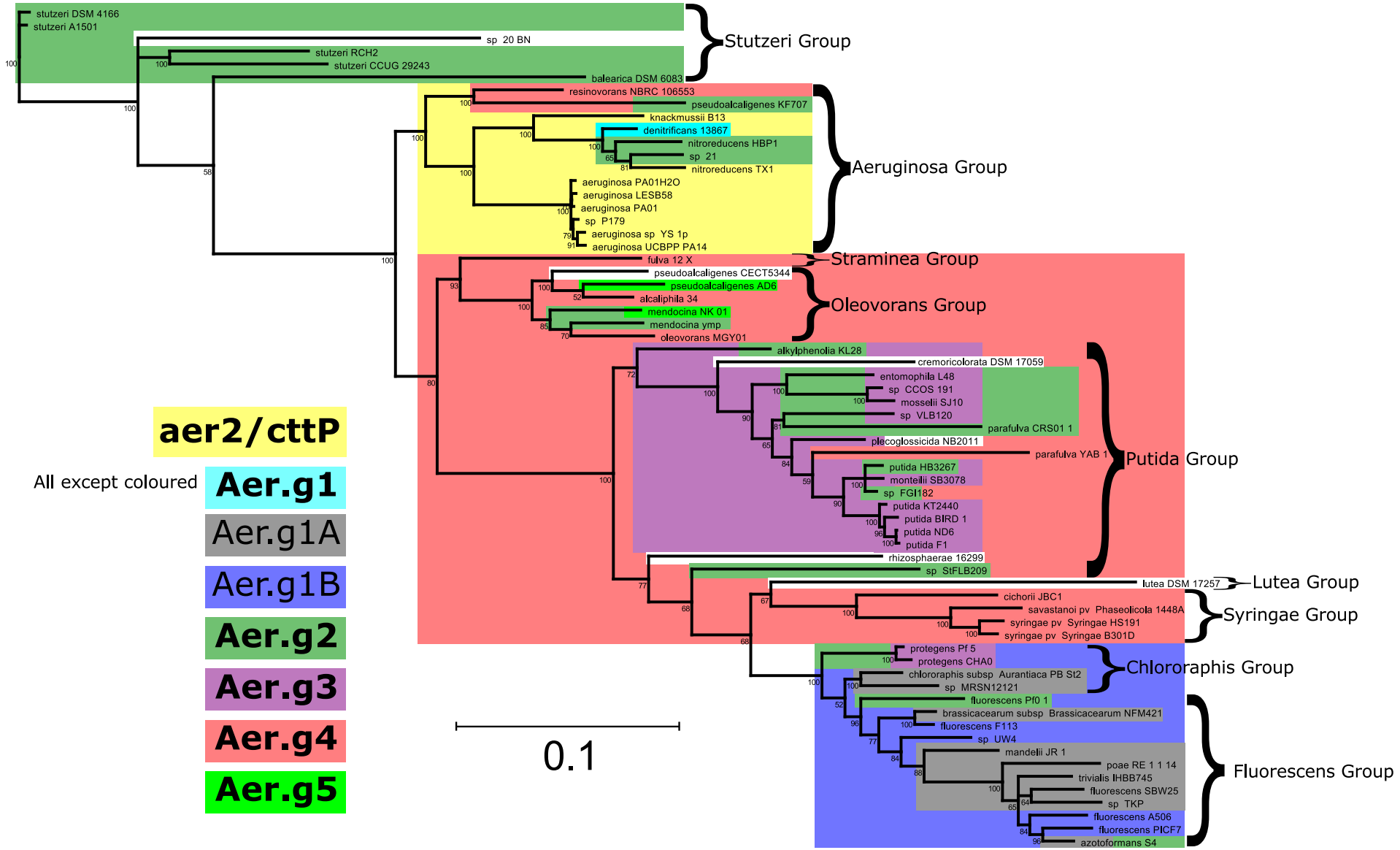


Figure 2: Unrooted maximum likelihood consensus tree showing phylogenetic relationship between *Pseudomonas* species based on concatenated *gyrB/rpoB/rpoD* nucleotide sequences. Colours indicate presence of *aer* homologs from each group as well as *aer-2/cttP*. Split colours and overlays indicate the presence of multiple homologs. All species have Aer.g1, except P. denitrificans 13867 (cyan). Numbers along branches indicate bootstrap support values from 100 replicates. Scale bar indicates average number of nucleotide changes per position.

175 Relationship between *Pseudomonas* and Aer Phylogenies

176 To get a better picture of how the phylogenetic relationships of Aer were influenced by
177 the intragenus phylogeny of *Pseudomonas*, a species phylogeny was generated (Figure 2). This
178 phylogeny was generated using concatenated alignments of *gyrB*, *rpoB*, and *rpoD* and matched
179 well with previously published *Pseudomonas* phylogenies (17, 23). Species clades were named
180 based on Gomila *et al.* (17) except for the *fluorescens* group as the present study did not include
181 the large number of species necessary to sub-divide this group. Examination of how the various
182 Aer subgroups, as well as Aer-2/CttP, mapped onto the genus phylogeny revealed their likely
183 evolutionary trajectories (Figure 2). The presence of *aer.g1* in all species, except *P. denitrificans*
184 indicates that it is a basal *Pseudomonas* gene. Both subdivisions of Aer.g1 occurred only in the
185 more derived *P. fluorescens* subgroup. Conversely, *aer.g2* is distributed throughout the genus
186 with no clear pattern. It is present in the more ancestral *P. stutzeri* group as well as the more
187 derived *P. fluorescens* subgroup but is not present in many species in between. *aer*.g3 is present,
188 except for *P. protegens*, exclusively in the *P. putida* group, though not all members of the clade
189 have it. *aer*.g4 is present mostly in the related *P. straminea, P. oleovorans, P. putida, P. lutea*
190 and *P. syringae sub*groups. Notably, it is not present in the *P. chlororaphis* or *P. fluorescens*
191 subgroups, despite these being more derived than the aforementioned groups. *aer.g4* was also
192 found in *P. resinovorans* and *P. pseudoalcaligenes* KF707, but not the rest of their *P. aeruginosa*
193 subgroup. *aer.g5* is only present in two species, both in the *P. oleovorans* subgroup. aer-*2* and
194 *cttP*, which were always found together, were exclusively present in the *P. aeruginosa* subgroup.

195 Genetic Features of *aer* Homologs

196 The grouping of Aer into 5 subgroups based on protein sequence information was
197 confirmed and expanded on by examining features of the underlying *aer* genes. The upstream

198 and downstream regions of the *aer* genes were inspected to determine whether the genomic
199 context was consistent across species (Figure 3). Additionally, the first two upstream and
200 downstream genes of each *aer* gene were identified (Supplementary Table 1). The frequency of
201 each of the associated genes and the general genomic context for each Aer homolog were thus
202 examined to identify the most commonly associated genes (summarized in Figure 3). Each
203 homolog was part of an apparently unique gene cluster: Aer.g1 with an aconitate hydratase (in
204 69% of species) and a CAAX aminoprotease (33%), Aer.g2 with a PAS-containing diguanylate
205 cyclase/phosphodiesterase (88%), Aer.g3 with a different PAS-containing diguanylate
206 cyclase/phosphodiesterase (88%) and Aer.g4 with a LysR-type transcriptional regulator (64%)
207 and a DTW-domain containing protein (74%). Beyond these, the second upstream and
208 downstream genes tended to vary more widely, though some upstream genes were clearly
209 conserved: Aer.g1, 23S rRNA methyl transferase (42%); Aer.g3, C4 dicarboxylate transporter
210 (63%); Aer.g4, agmatinase (56%).

211 Evidence for Horizontal Transfer of *aer* Homologs

212 While examining the genomic contexts of *aer* homologs, indicators of mobile genetic
213 elements were found nearby in many cases (Supplementary Table 2). Along with the varied
214 distribution of Aer homologs within the *Pseudomonas* genus, and the findings from the genus
215 phylogeny that distantly related species had similar complements of *aer* genes, this implied that
216 they may have been subject to horizontal transfer. This possibility was further examined by
217 generating tanglegrams comparing each of the *aer* subgroup phylogenies to corresponding genus
218 phylogenies (Supplementary Figures 6-9). Initial inspection of the phylogenetic tree identified 3
219 sequences that diverged from the expected species phylogeny. These 3 sequences
220 (WP_015271024.1, *P. putida* HB3267; WP_014754514.1, *P. putida* ND6; and WP_013791017.1,

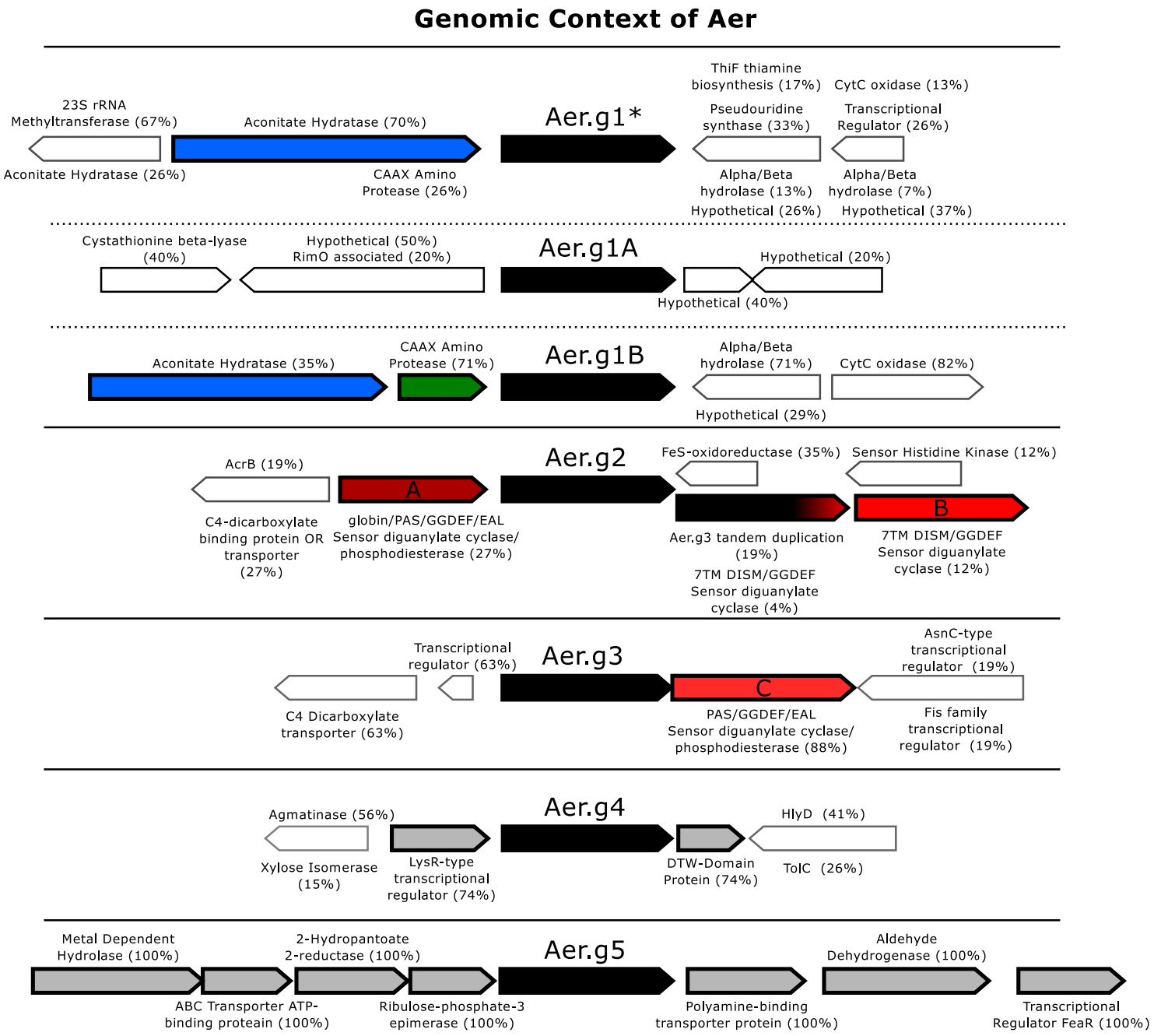


Figure 3: Frequency of occurrence and orientation of genes upstream and downstream from *aer* homologs from select *Pseudomonas* species. Genes with thick outlines are expected to be part of the same gene cluster based on their orientation and high frequency of occurrence with *aer*. Genes with noteworthy functions are coloured, and are discussed in the text, such as the 3 different diguanylate cyclases/phosphodiesterases (A, dark red; B, red; C, light red). Numbers beside gene functions indicate the frequency that they were found within each group or sub group. Gene lengths are approximate, *aer* is about 1.5kb long. Aer.g1* indicates Aer.g1 sequences not including Aer.g1A and Aer.g1B.

P. fulva 12-X) were each found clustered with sequences from unrelated species whereas other Aer sequences were consistently clustered with sequences from closely related species. Of these 3 genes only the *aer.g3* gene from *P. putida* ND6 was associated with mobile elements. Mobile elements were found within 5 Kbp of *aer* homolog genes (Supplementary Table 2). Inverted repeats were most common (20%), followed by transposases (4%) and integrases (1%). Aer.g5 was never associated with mobile elements, Aer.g1 was associated least frequently, then Aer.g3, then Aer.g2 and Aer.g4 had the highest frequency of association with mobile elements. When mobile elements were found, they were often located immediately up and/or downstream of the *aer* homolog and its associated gene(s) described in the above section. Tanglegrams were used to identify instances where the *gyrB*/*rpoB*/*rpoD* based species phylogeny did not match with the *aer* subgroup phylogenies (Supplementary Figures 6-9). Three instances for Aer.g2 and one for Aer.g3 were identified where the *aer* gene had likely been transferred horizontally. Probable instances of horizontal transfer were also identified from mapping the *aer* subgroups onto the genus phylogeny (Figure 2). *P. resinovorans* and *P. pseudoalcaligenes* KF707 have Aer.g4 despite being more distantly related to the other strains that possess this homolog. Also, both *P. protegens* strains have Aer.g3 despite being more distantly related to the *P. putida* subgroup strains that only have this homolog.

Amino Acid Sequence Comparison of Groups

As genomic analysis of the Aer subgroups confirmed the AA sequence-based groupings from the SHMR analysis, further sequence analysis was pursued to uncover sequence features that distinguished the various Aer groups. First, the domain architecture from all sequences were compared (Supplementary Figure 10). Domain assignments were obtained by submitting all sequences to SMART (24). All sequences were similar to the expected domain architecture of

Aer from *P. aeruginosa* PA01, consisting of a N-terminal Per/Arnt/Sim (PAS) domain, transmembrane helices and then the cytoplasmic kinase control (called MA for methyl-accepting in SMART) region which includes the CheW/CheA interface. Most Aer.g2 sequences also had a HAMP domain between the transmembrane region and MA domain, similar to Aer from *E. coli* (25). To further identify regions of the protein that are specific to the subgroups, results from the SHMR analysis for each inter-group comparison were plotted against the entire length of the Aer protein (Figure 4) and compared to the overall conservation at each AA position (Supplementary Figure 11). As expected, the CheW interface region of the MCP-signal domain and the PAS ligand binding domain were the most conserved regions, and also did not differ between groups, though the PAS domain was less conserved. Inspection of the SHMR scores for each intergroup comparison showed that the regions determining the specificity of each group were the same for each Aer group (Figure 4), except for Aer.g2 which had a group-specific region immediately N-terminal to the PAS domain. Sequences from this group also had variable length C-terminal extensions. The most notable group-specific region was in between the transmembrane helices and the beginning of the kinase control (MA) domain. As this is the location of the HAMP domain in *E. coli* Aer, this region was examined more closely.

Group-Specific Domain Differences

First, the CheW/CheA interface region was examined more closely as its subdomains have been clearly defined (26). In bacterial chemoreceptors this domain consists of pairs of heptads, separated into three subdomains. Central to the domain is a glutamate residue which denotes the 'zero' position from which paired heptads are counted outwards until the C-terminus on one side and the end of the domain on the other. As there were 20 heptads to the C-terminus in all sequences, except Aer.g2 which had variable length extensions, all Aer from *Pseudomonas*

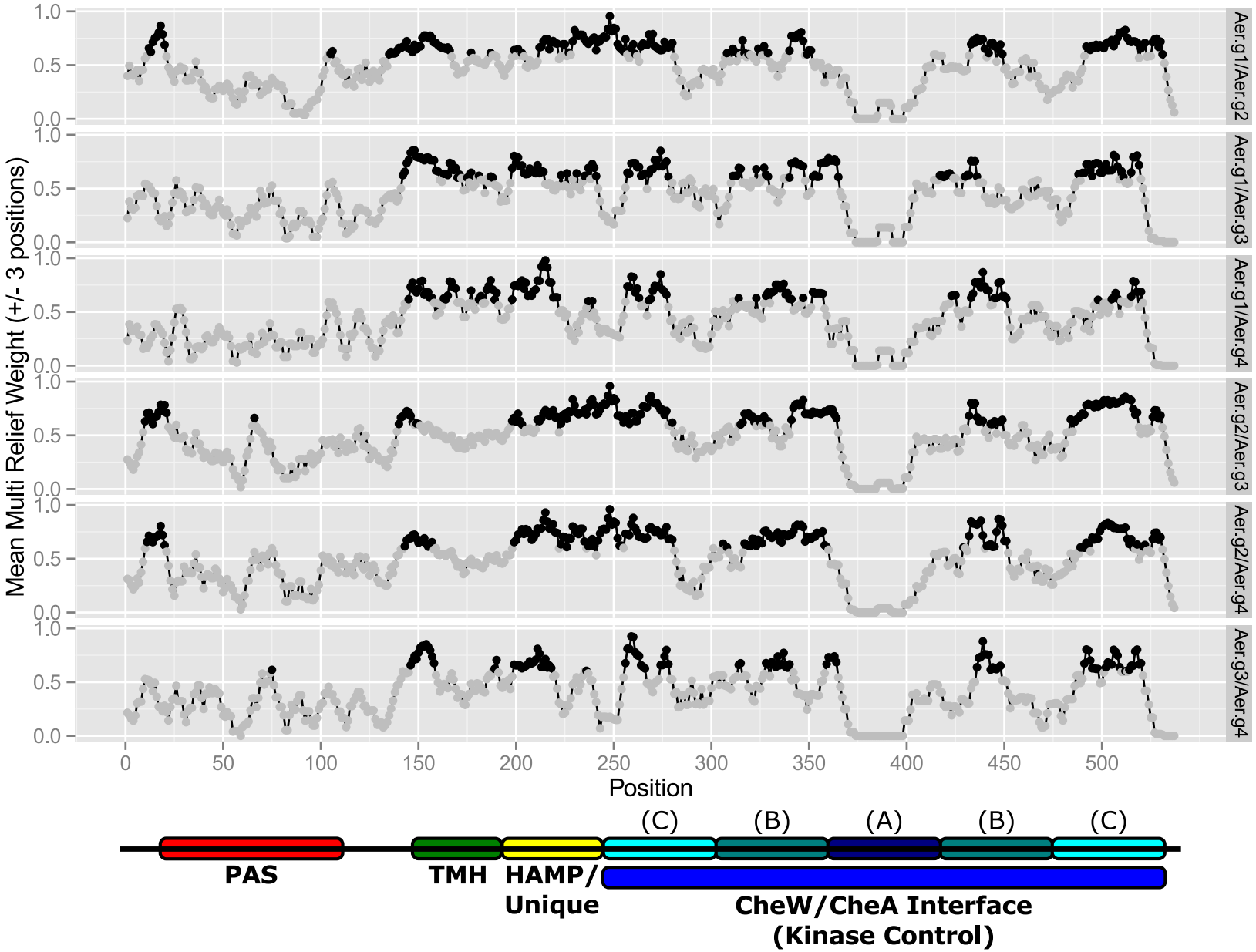


Figure 4: *Pseudomonas* Aer domain architecture and multi-relief scores showing group specificity of amino acid positions across the entire length of Aer for all intergroup comparisons. Black points indicate positions with a score above 0.6, grey below. This cutoff indicates positions that are conserved within groups but divergent between. Domains are PAS (Pern/Art/Sin) ligand binding (red), TMH (transmembrane helices, green), HAMP/Unique (in Aer.g2 HAMP, unique domains in other groups, yellow), CheW/CheA Interface also called Kinase Control (subdivided into A, signaling, dark blue; B, flexible bundle, teal; C, methylation, cyan).

fall in the 40H class (26). The three sub-domains (from the center outwards), signaling, flexible bundle, and methylation were marked and used to compare the conservation (Supplementary Figure 12) and group-specificity (Supplementary Figure 13) of the various Aer groups. Conservation showed that corresponding heptads of the more central methylation subdomain were better conserved than the flexible bundle subdomain. This corresponded with indications of group-specific features in the flexible bundle subdomain and in the outer heptads (furthest from the signaling subdomain) of the methylation subdomain.

The PAS and HAMP domains were also compared between all groups. The PAS domains from all Aer groups were not notably different and all had the expected features of a PAS domain (Supplementary Figure 14). Conversely, the cytoplasmic region between the transmembrane helices and kinase control domain, which contains a HAMP domain in *E. coli* Aer, had unique features in each Aer group (Figure 5). Only Aer.g2 had characteristic AAs of a HAMP domain, though all groups could be anchored by a conserved aspartate and glycine (which differed in absolute position between groups). The only other notable conservation was the final seven AAs of the domain, which were only conserved between Aer.g1, Aer.g3 and Aer.g4. In this heptad with the motif SZEARL(K/Q), only the Z was present in Aer.g2. With regards to the subdivision of Aer.g1, both the PAS and HAMP/'unique' domains supported keeping all sequences as a single group (Supplementary Figures 14 and 15). Aer.g2 also differed from the other three groups at its C-terminus as the final heptad of the kinase control domain differed and its members feature an up to 9 AA extension (Supplementary Figure 16).

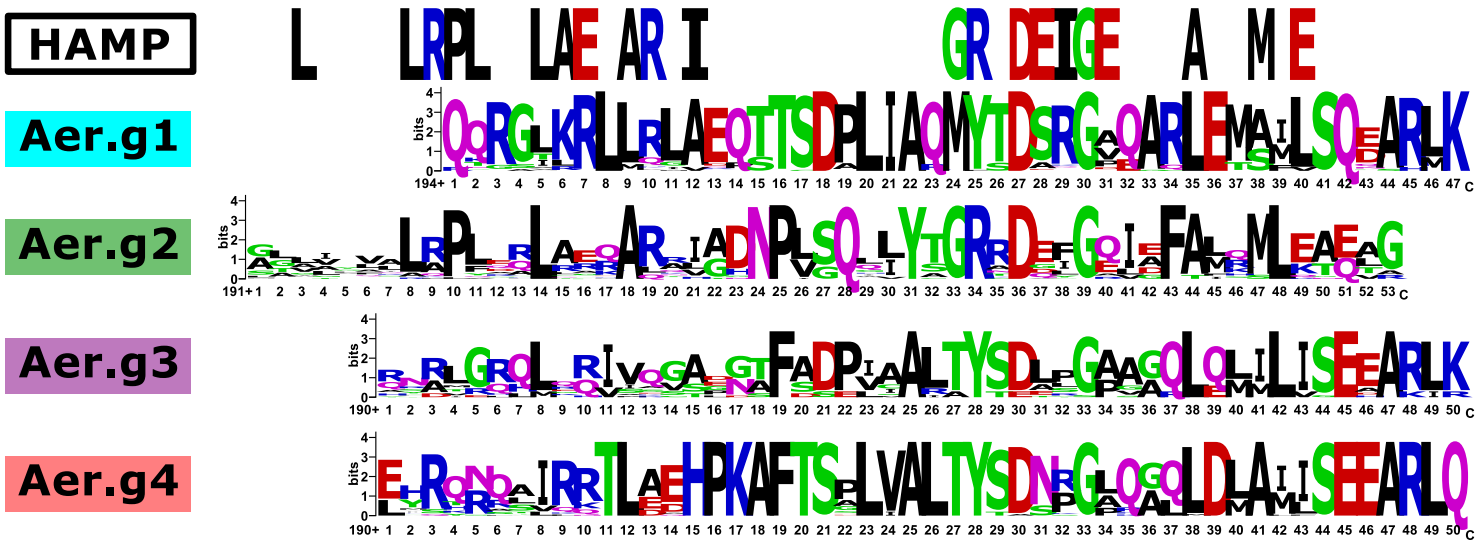


Figure 5: Weblogos of Aer group specific region in between end of transmembrane domain and start of cytoplasmic domain heptads. Start of cytoplasmic domain was based on the number of matching heptads on either side of the central glutamate residue (Aer.g1, 388; Aer.g2, 394; Aer.g3, 394; Aer.g4, 391). The end of the transmembrane domain was determined by submitting the consensus sequence of each group to THMMR. Weblogos were aligned based on the shared aspartate residue (Aer.g1, 221; Aer.g2, 227; Aer.g3, 220; Aer.g4, 220). Characteristic HAMP domain residues were obtained from the SMART database.

Phenotypic Results

To investigate the functions of Aer homologs, CttP and Aer-2, and to see whether they were linked, deletion mutants were constructed in the one species of *Pseudomonas* that possessed three Aer homologs, Aer-2 and CttP: *P. pseudoalcaligenes* KF707. Single and combination deletions were constructed using 2-step homologous recombination, the suicide vector pG19II and SacB sucrose counter-selection (27). To test energy-taxis, "soft agar swarm plates" were used (28). As swarming motility is NOT being tested in this assay, we called them "energy-taxis swim plates" instead of the term "soft agar swarm plates" in previous studies which is misleading. In this assay, a small amount of cells were inoculated into a plate by stabbing a needle covered in liquid culture into an agar plate containing a low (0.3%) percentage of agar and a high amount of carbon source (50mM) in minimal salts. As the initial inoculum divides and consumes the carbon source at the point of inoculation, daughter cells will swim outwards, producing a ring as energy-tactic cells seek out a better place to grow. Non-chemotactic strains will not be able to extend beyond the point of inoculation but non-energy tactic cells will produce a smaller ring than the wild-type (Supplementary Figures 17 and 18). To ensure that energy-taxis was being compared between the different mutants, the results were compared to a non-chemotactic CheA::KmR insertional inactivation mutant (29), and the chemotactic ability of all strains was also confirmed in classical chemotaxis swim plates (Supplementary Figure 19). Comparison of these metabolism independent results to the metabolism dependent results of the energy-taxis swim plates allowed for differences in energy-taxis capabilities to be observed between strains. All 24 mutants generated in this study, along with the wild-type and CheA control were tested in triplicate in media containing pyruvate (Figure 6) or succinate (Supplementary Figure 20). Ring diameters were measured from the

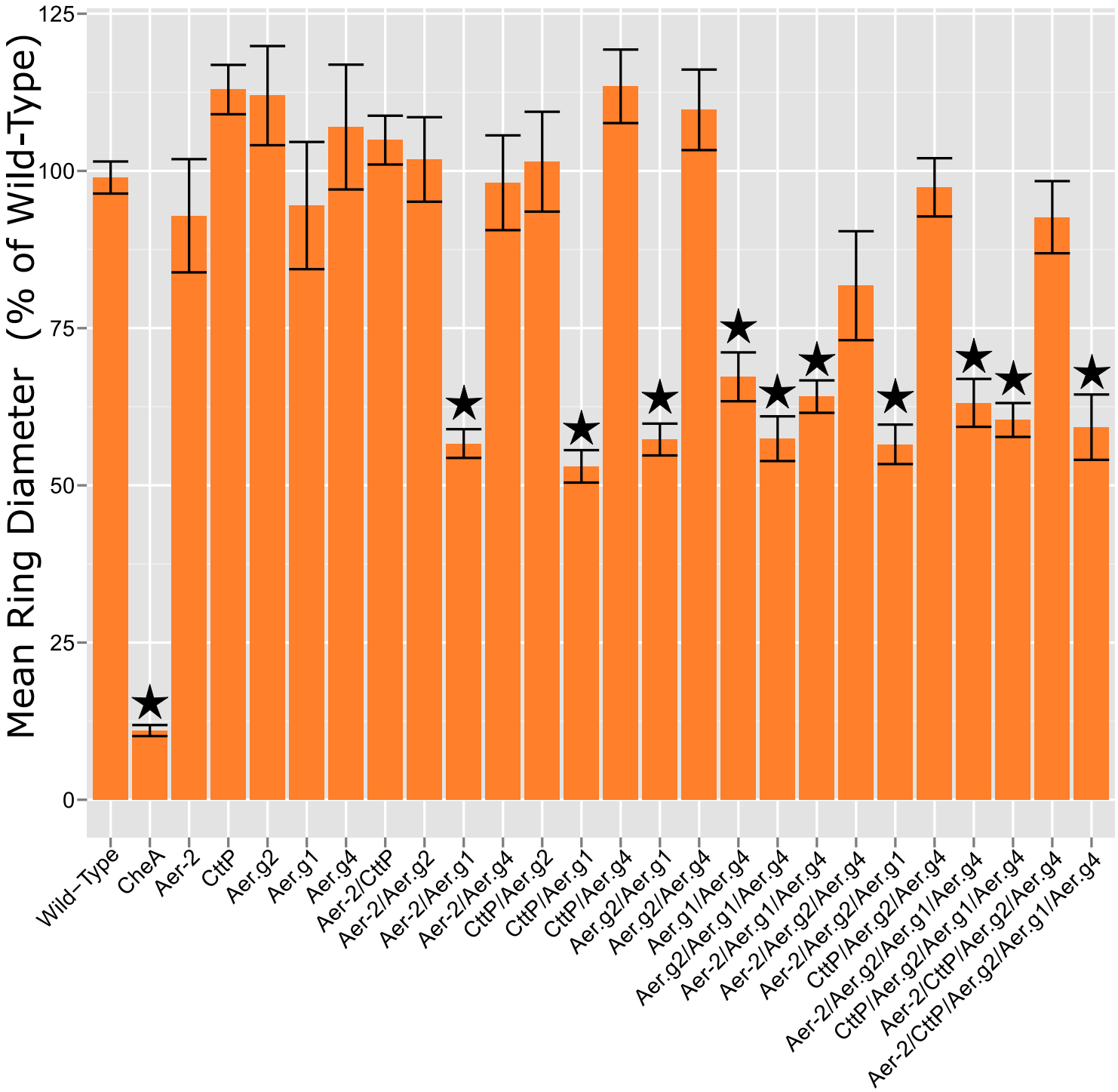


Figure 6: Normalized energy-taxis growth diameters in 50mM pyruvate of strains of *P. pseudoalcaligenes* KF707 with deletions of *aer2, cttP* and *aer* homologs. Bars indicate the average growth diameter, normalized to the wild-type, at both 24h and 48h from at least 3 experimental replicates. Wild-type strains were normalized to the mean of technical replicates within each experiment. Error bars indicate standard error. Stars indicate significant differences from the wild-type based on Tukey's Honest Significant Differences test with a confidence value of 0.95.

313 furthest point reached from the inoculation centre (Supplementary Figure 18), normalized to the
314 wild-type, then tested for differences using Tukey's Honest Significant Differences test
315 (Supplementary Table 3).

316 The diameter of the chemotactic negative CheA mutant was about 10% of the wild-type,
317 whereas energy-taxis negative mutants with diameters about 60% smaller were significantly
318 different from the wild-type (Figure 6). All the singular deletions of Aer.g1, Aer.g2, Aer.g4, Aer-
319 2 and CttP had no effect on energy-taxis; only strains that had Aer.g1 deleted in combination
320 with at least one other receptor showed changes in their energy-taxis phenotype. Conversely,
321 deletion of any combination of these receptors, including the quadruple Aer-
322 2/CttP/Aer.g2/Aer.g4 mutant did not adversely affect energy-taxis. The Aer-2/Aer.g2/Aer.g4
323 triple mutant had a diminished diameter, about 80% of the wild-type, but this was not significant.
324 Results from plates containing succinate (Supplementary Figure 20) were similar to those
325 obtained using pyruvate, but were more variable. The observed differences in swim diameters
326 were not due to differences in growth rate as the rate of growth of the swim diameters was
327 similar between all strains (Supplementary Figure 21, Supplementary Table 4) as well as their
328 growth in liquid media (Supplementary Figure 22). All strains were also tested for chemotaxis
329 towards pyruvate, glucose (Supplementary Figure 19) and succinate (data not shown) using
330 traditional swim plate assays, and all strains except the CheA::KmR mutant were able to swim
331 towards pyruvate, succinate and glucose. As Aer.g1 played a pivotal role in energy-taxis the
332 wild-type and quintuple Aer-2/CttP/Aer.g1/Aer.g3/Aer.g4 mutant were complemented with
333 Aer.g1 on a plasmid which showed the expected increase and recovery, respectively, of energy-
334 taxis in succinate plates (Supplementary Figure 23, Supplementary Table 5). Together these

335 experiments showed that it was only the combined deletion of Aer.g1 with any of the other tested
336 receptors that caused a reduction of energy-taxis.
337
338 **Discussion**
339 In *Pseudomonas* the chemoreceptor Aer has been characterized as a receptor for energy-
340 taxis. *P. aeruginosa* has a single Aer chemoreceptor whereas *P. putida* has three homologs. Here
341 we used a bioinformatics analysis of Aer protein sequences from throughout the *Pseudomonas*
342 genus to determine the relationship between Aer homologs, their specific features, and generate a
343 phylogenetically-based naming for these homologs to alleviate the problematic name given to
344 'Aer2'. Additionally, in *P. pseudoalcaligenes* KF707, the one species that possessed three Aer
345 homologs, Aer-2 and the ostensible tetrachloroethylene receptor CttP, deletional mutagenesis
346 was used to demonstrate that the Aer.g1 was key but not essential for energy-taxis.
347 Protein sequence and genetics classify Aer homologs into 5 groups
348 Aer has previously been described in *P. putida* and *P. aeruginosa* (8, 30–32), though the
349 two species differed in the number of apparent homologs indicating there is interspecies
350 variation. Here we constructed a protein sequence based phylogeny which separated Aer into 5
351 different subgroups. All groups varied in the same regions of the protein, except Aer.g2 which
352 had a unique region at the C terminus, which does not resemble the C-terminal cheR interface
353 pentapeptide of Aer-2 (33). The most conserved regions included the PAS ligand binding domain
354 and CheW/CheA interface region which confer the necessary functions of FAD cofactor
355 binding/signal reception (34) and signal transduction, respectively. The conserved regions in the
356 signaling region were better conserved than the predicted PAS domain, though these domains are
357 known to differ at the sequence level (34). As the PAS domains were conserved, this suggests

that all Aer subgroup members respond to the same signal. In all groups, the N-terminal heptad 13 methylation sites (35) were also conserved, implying that Aer is modulated via methylation similarly to other chemoreceptors. Interestingly, it was the cytoplasmic region between the transmembrane helices and kinase control domain that was unique to each subgroup (Figure 5). Aer.g2 had a HAMP domain in this location, as in *E. coli* Aer (36). In this organism, interactions between the PAS and HAMP domains are essential for the function of the protein as the HAMP domain transfers the signal through the protein (25). Aer.g2 likely functions in the same way but the other four Aer groups do not have this HAMP domain, suggesting that they function differently. While there were a few positions conserved between the groups, such as the aspartate used to anchor this portion of the alignment (Figure 5), the most important defining residues (37) were only present in Aer.g2. HAMP domains are also expected to be found here in the majority of MCPs (2) so this indicates that most Aer homologs in *Pseudomonas* differ substantially from other MCPs.

Protein sequence based group assignments were supported by the genomic context of the corresponding genes; there tended to be the same genes in the immediate vicinity of the *aer* genes from each group, though this also supported the existence of additional groups. The two branches of Aer.g1 which were initially classified as their own groups (Figure 1) did not have the same up and downstream genes as the rest of Aer.g1, however including these sequences within Aer.g1 was supported by the SHMR analysis. These sequences represent the only duplications of Aer.g1 and only occurred in species which had no other Aer homologs (except *P. azotoformans* S4). As the protein sequences do not differ as much as the other groups, and the genomic context within this group differed, this may represent a recent duplication event within these lineages.

Phenotypic characterization of the two Aer receptors in these species could be particularly enlightening to understanding why only some *Pseudomonas* species have extra *aer* genes.

Aer.g1 is the most prevalent homolog

Aer from *E. coli* has been well-studied, so Aer from *E. coli* K12 was used as an outgroup for the *Pseudomonas* Aer phylogeny, even though *E. coli* is not closely related to *Pseudomonas* (38). *E. coli* Aer was distantly related to all *Pseudomonas* Aer sequences and it was positioned between Aer.g1 and Aer.g2 on the phylogenetic tree (Supplementary Figure 1). As such, which of these groups was ancestral could not be determined from the protein sequence phylogeny alone. Aer.g1 was the most prevalent as all the *Pseudomonas* species included in this study had at least one homolog except *P. denitrificans* ATCC13867. As this species did not have any Aer homologs, this means that Aer is not part of the *Pseudomonas* core genome, though it is likely ancestral and was lost in *P. denitrificans*. Unlike Aer.g1, Aer.g2 was not present in all *Pseudomonas* species, but it was present in species from most subgroups. Together with the *E. coli* rooted phylogeny this suggests that both Aer.g1 and Aer.g2 were present in the last common ancestor of all *Pseudomonads* but only Aer.g2 was present in the last common ancestor of *E. coli* and *Pseudomonas*. This would imply that Aer.g2 is basal to the gamma proteobacteria. Indeed, preliminary investigations suggest homologs are present in diverse lineages such as *Marinobacter, Shewanella, Ralstonia,* and *Vibrio.*

Almost all *aer.g1* genes were adjacent to an aconitase gene (Figure 3), indicating that they may be co-regulated or co-transcribed. In *Helicobacter pylori*, TlpD, which controls tactic behaviour in low-energy conditions, has been found to interact with aconitase (39), possibly indicating some common connection between aconitase and energy-taxis receptors. A gene encoding a predicted CAAX amino protease was found adjacent to one third of Aer.g1 homologs.

403 This type of membrane-bound protease aids in prenylating proteins to ease their membrane
404 localization (40). The Aer.g1 proteins whose genes were located beside these predicted proteases
405 did not have the expected CAAX motif and there were no obvious features at the sequence level
406 that distinguished these sequences from other Aer homologs indicating that the genomically co-
407 localized CAAX protease may not be involved in their functioning.

408 Other Aer groups may have unique functions

409 As the 5 previously characterized Aer receptors are Aer.g1 homologs with the same
410 energy-taxis function, it is possible that the other homolog groups may have a related, but
411 different or more specific function. The three Aer homologs of *P. putida* KT2440 were found to
412 be differentially expressed (9), supporting this hypothesis. In *E. coli* Aer is also a thermosensor
413 (41) indicating another possible role, or they could aid in tuning the energy-taxis response
414 similarly to how Tsr and Tar from *E. coli* enable taxis towards the ideal pH from both higher and
415 lower initial pHs (42). The function of the additional Aer homologs could also be related to
416 biofilm formation as Aer.g2 and Aer.g3 were each associated with their own PAS-domain
417 containing diguanylate cyclase/phosphodiesterase. Though Aer is expected to only act as a Che1
418 chemosensor (35), the recent revelation that not all chemosensors mediate chemotaxis (43), and
419 are instead involved in so-called 'alternative cellular functions' such as cyclic di-GMP signaling,
420 makes this frequent association of phosphodiesterase/diguanylate cyclases with Aer.g2 and
421 Aer.g3 noteworthy. Deletion of the diguanylate cyclase/phosphodiesterase associated with
422 Aer.g3 in *P. putida* KT2440 caused a general defect in motility (9) and was later shown to be a
423 bi-functional cyclic di-GMP phosphodiesterase and diguanylate cyclase (44). These two
424 homolog groups may thus be involved in energy-sensing behaviour in biofilms as c-di-GMP is
425 an important regulator in the transition from planktonic to sessile growth modes. In *P.*

426 *aeruginosa*, increased levels of cyclic di-GMP decreases the frequency of flagellar motor
427 switching through MapZ as it inhibits CheR1 from methylating some MCPs (45). As this shows
428 a direct connection between chemotaxis and cyclic di-GMP, understanding the function of the
429 Aer homolog-associated cyclase/phosphodiesterases will be of interest.

430 Possibility of horizontal transfer of *aer* homologs

431 Chemoreceptors have been posited to be subject to horizontal gene transfer (46), but to
432 our knowledge this has only been demonstrated in plasmids of *E. coli/Shigella* (47). Here we
433 presented multiple lines of evidence that suggest genes from the *aer* group have been
434 horizontally transferred in the chromosome of *Pseudomonas* species. In the Aer phylogeny, three
435 homologs from *P. putida* HB3267, ND6 and *P. fulva* 12-X each clustered beside Aer sequences
436 from unrelated species indicating they may have been obtained from those species.
437 Incongruencies between the Aer phylogeny and genus phylogeny were also found using
438 tanglegrams. These figures highlight other instances where *aer* homolog sequences were more
439 closely related than the species hosting these genes, which could indicate non-vertical
440 inheritance. Additionally, we noted several cases where strains of the same species differed in
441 which Aer homologs they possessed. For example, of the three *P. pseudoalcaligenes* strains
442 examined here, each one had a drastically different complement of genes (CECT5344, *aer.g1*;
443 AD6, *aer.g1/aer.g4/aer.g5*; KF707, *aer.g1/aer.g2/aer.g4/aer-2/cttP*).

444 The observed variation in the distribution of Aer homologs is unsurprising as even within
445 the *P. aeruginosa* pan-genome from 7 strains there are 2,000 accessory genes compared to 5,000
446 core genes (48). The fact that *aer.g1* almost qualifies as part of the genus core genome (*P.*
447 *dentrificans* being the only outlier) indicates that is a very useful gene in the varied environments
448 inhabited by *Pseudomonas*. The existence of the various homologs speaks to their utility as they

each may confer a more specific function. As many *Pseudomonas* species live in the soil and rhizosphere, which promotes horizontal gene transfer (49), there would be ample opportunity for these *aer* homolog genes to spread. Interestingly they appear to be restricted to *Pseudomonas* as reverse BLAST searches consistently returned *Pseudomonas* sequences as the closest hits.

Genes influencing energy-taxis in *P. pseudoalcaligenes* KF707

P. pseudoalcaligenes KF707 was the only species that had three Aer homologs as well as Aer-2 and CttP. We have previously examined its chemotactic ability towards biphenyl (50), and how its multiple *cheA* genes affect biofilm formation (29) and swimming motility (51). Aer from *P. aeruginosa* PA01 (8), *P. putida* F1 (11) and *P. putida* KT2440 (9) have been previously characterized but are all part of the Aer.g1 homolog group. The functional Aer homolog from *P. putida* PRS2000 (10) was not included here as this species' genome has not been fully sequenced, though based on its sequence it is also an Aer.g1 homolog. The receptors from *P. putida* F1 and KT2440 were both named 'Aer2' but the phylogeny presented here indicates they belong to the Aer.g1 homolog group and so should be renamed to remove the '2'. In the present study, all three *aer* homologs in *P. pseudoalcaligenes* KF707, representing *aer.g1*, *aer.g2* and *aer.g4*, as well as *aer-2* and *cttP* were deleted individually and in combination. The energy-taxis phenotypes of these mutants were compared based on their diameters in soft agar energy taxis plates. As the growth rate of all the strains in these plates, and in liquid culture did not differ, as well as their chemotactic ability in swim plates, the observed differences were taken as indication that the perturbed genes influenced energy-taxis in some way.

No single deletions affected the energy-tactic ability of *P. pseudoalcaligenes* KF707. This differs from *P. putida* KT2440 which became energy-taxis negative when only its Aer.g1 homolog was inactivated (31). Both species have Aer.g4, but KF707 has Aer.g2 and KT2440

Aer.g3. In KF707, co-deletion of either *aer.g4* or *aer.g2* with *aer.g1* caused a reduction of energy-taxis, but deletion of *aer.g2* and *aer.g4* together had no effect. The three Aer homologs in KT2440 are differentially expressed and thus were not truly redundant (31), but in KF707 its Aer homologs appear to be at least partially redundant. In *Ralstonia solanacearum* both of its "Aer1" and "Aer2" receptors were necessary for aerotaxis and restored energy-taxis in a deficient *E. coli* mutant (52). As these receptors appear to be homologous to Aer from *Pseudomonas* and *E. coli*, it was likely energy-taxis enabling the swimming of *R. solanacearum* towards oxygen. Interestingly these two receptors appear not to be redundant as inactivation of either one caused a defect. In *Vibrio cholerae*, the number of Aer homologs also differs between species. The El tor biotype has three, compared to two in strain O395N1 (53). The additional Aer (called Aer-1) in the El tor biotype is most similar to Aer.g2, though all the sequences from *Vibrio* and *Ralstonia* appear to be quite different (identity below 50% on <90% coverage) than *Pseudomonas* sequences making it difficult to categorize these homologs according to the naming scheme developed here for *Pseudomonas.* Only one homolog in *V. cholera* was found to affect energy-taxis, leaving the exact function of the other homologs unknown, similar to *Pseudomonas.* Of these five receptors from species other than *Pseudomonas* only one has an obvious HAMP domain, implying that the HAMP domain in *E. coli* Aer is less common than other forms of Aer in gamma proteobacteria.

Aer and Aer-2 were first characterized as aerotaxis receptors in *P. aeruginosa* PA01 (8). In *P. putida* KT2440, which does not have Aer-2, its Aer.g1 homolog was found to mediate taxis towards oxygen (9). Aer-2 has been shown to bind oxygen (12), and its overexpression in *P. aeruginosa* and *E. coli* abolished chemotaxis (54) but its function as an aerotaxis receptor remains inconclusive. Here we showed that single deletion of *aer-2* did not affect energy-taxis,

but co-deletion with *aer.g1* resulted in a similar reduction in energy-taxis as co-deletion with either of the other *aer* homologs. The Aer-2 protein has a C-terminal extension that specifically allows CheR2, and only CheR2, to methylate it (33), and is required for the complexation of other Che2 proteins (13), making it appear that the entire Che2 pathway exists to transduce the signal sensed by Aer-2. This was recently confirmed using sequence analysis of all *P. aeruginosa* PAO1 chemoreceptors (35). Our results were thus surprising as no connection was expected between Aer-2 and the che1 system, which transduces the signal of Aer, though the functions of Aer and Aer-2 in aerotaxis were linked in their initial characterization (8). As their phenotypic outputs may overlap, *P. denitrificans* may be an ideal organism to study Aer-2, as it has no Aer homologs.

The *che2* gene cluster is preceded by *cttP* so we investigated the possibility of its involvement in energy-taxis as well. In *P. aeruginosa* PA01, CttP was characterized as a receptor for positive chemotaxis towards tetra/tri-chloroethylene (PCE/TCE) (14). As this organism does not perform reductive dehalogenation, intentional taxis towards these compounds is puzzling. In *E. coli*, the receptors Tar and Tsr can mediate attractant and repellent responses to phenol through its direct interaction with the transmembrane helices or cytoplasmic HAMP signaling domains (15). We thus hypothesized that a similar interaction may have mediated the observed PCE/TCE chemotaxis and that CttP actually has a different function. CttP is only found in organisms with Aer-2, and the gene is located immediately upstream of the *che2* gene cluster. Here we showed that co-deletion of *cttP* and a*er.g1* in *P. pseudoalcaligenes* KF707 reduced energy-taxis comparable to co-deletion of *aer.g2* or *aer.g4*. CttP has an unusual domain architecture compared to other MCPs as it has no predicted ligand-binding domain and a single N-terminal predicted transmembrane domain (55). It does have the expected CheW/CheA

518 interface domain, but also a long (>100 AA) C-terminal extension not seen in most MCPs. It
519 could be possible that CttP does not directly sense any signal, but interacts with MCP receptor
520 complexes and/or influences their interaction with the Che complex. Overexpression of *cttP*
521 abolished all chemotaxis in both *P. aeruginosa* and *E. coli* (54), possibly indicating that its
522 function is in modulating signaling from the che1 system.

523 Despite being the first gene in the *che2* gene cluster, *cttP* is involved only with *che1*
524 genes as the *che2* system is dedicated to processing the signal sensed by *aer-2* (35). It is thus
525 puzzling that *cttP* was always found in species that have *aer-2*, and that deletion of *aer-2* in
526 combination with *aer.g1* diminished energy-taxis in *P. pseudoalcaligenes* KF707. This could be
527 due to cross-talk between the systems as the exact output of the *che2* system is not fully known,
528 though it is known to affect chemotaxis. The observed phenotype could have also been due to
529 overlapping phenotypic output between energy-taxis and aerotaxis as *aer-2* certainly binds
530 oxygen but its cellular function is not conclusively known (56).

531

532 **Conclusions**

533 In *Pseudomonas*, Aer is not a single receptor for energy-taxis, but a family of receptors.
534 The various homologs differ mostly in the region between the PAS ligand binding domain and
535 methyl-accepting CheW/CheA interface domain, and most do not have a HAMP domain. The
536 number of Aer homologs varies widely between species and they have likely been horizontally
537 transferred. The most prevalent homolog in the *Pseudomonas* genus, here named Aer.g1,
538 mediates energy-taxis in *P. aeruginosa* and *P. putida*, but in *P. pseudoalcaligenes* KF707 Aer.g2
539 and Aer.g4, can also influence energy-taxis. Aer-2 and CttP also appear to have some effect.

These findings indicate that Aer and energy-taxis in *Pseudomonas* are more variable and complicated than the single receptor found in the archetypal species *P. aeruginosa.*

Materials and Methods

Protein Sequences

All sequences were obtained using NCBI databases and tools (57). BLAST searches on the NCBI website (18) (pBLAST for draft and completed genomes, tBLASTn for whole-genome shotgun genomes) were performed using *P. aeruginosa* PA01 Aer (NP_0250252.1) as a query sequence. From each species, all hits with >95% sequence coverage, no matter how low the sequence identity, were selected for inclusion. BLAST did not return any results with coverage values between 67% and 95%, indicating that all included sequences were likely to truly be Aer sequences. Expect values were always below $1x10^{-100}$. Four sequences were removed as they were redundant entries resulting from incorrect start site annotations resulting in two proteins with the same C-terminus but slightly different N-termini. Sequence accession numbers were thus obtained from the international nucleotide sequence database collaboration (INSDC) (58). NCBI Entrez was used to obtain FASTA formatted sequences which were then compiled into a single file.

Amino Acid Alignment

Full length sequences were aligned using COBALT (on the NCBI website) (19) as this algorithm works well when only part of the sequence is well conserved, as in MCPs. The default settings were used as adjustment did not improve the alignment noticeably. The generated alignment had numerous gaps, which were manually removed by adjusting the alignment in Jalview 2.10 (59). The alignment was then finalized by re-aligning using MUSCLE (as

563 implemented in Jalview, default settings) (60), resulting in a removal of gaps without
564 misaligning the sequences. Names were cleaned up using a custom script in R to reduce the
565 names to just the species, strain and accession number in a presentable fashion (61). An
566 additional alignment including Aer from *Escherichia coli* was also generated in the same fashion.
567 Both alignments were used as input to generate maximum likelihood phylogenies using PHYML
568 (20), as implemented by the South of France Bioinformatics platform (21). PHYML options:
569 Amino-Acids, 100 bootstrap replicates, JTT amino-acid substitution model, BEST tree topology
570 search operation, tree topology, branch length and rate parameters were optimized. Consense
571 was then used to generate a consensus tree from the bootstrap replicates using the majority rule
572 (extended), and only treating the tree as rooted when the *E. coli* outgroup was included.

573 Sequence Harmony and Multi-Relief Analysis to Determine Groups

574 The multi-Harmony server was used to apply sequence harmony and multi-relief
575 (SHMR) to validate groupings made based on the ML tree grouping and alignment. Groups were
576 manually decided based on the tree topology and corresponding alignment, initially making for 7
577 groups. These seven groups were compared in pair-wise fashion using SHMR (22). Thus for
578 each pair of groups a score was calculated for each amino acid position that indicated how
579 conserved it was within groups and how divergent it was between groups. A score of 1 indicates
580 perfect conservation within and perfect divergence between groups. Empirical cumulative
581 distribution functions (ECDF) were plotted for each comparison and the percent of AAs above
582 the 0.8 cutoff recommended by the SHMR authors was determined. As most comparisons
583 resulted in 25% of AAs reaching this threshold, the two pairs of groups that, when compared,
584 only had 10% of AAs above the cutoff, were deemed incorrect group assignments. These groups
585 were merged and the process repeated to produce 5 groups that all had ~25% of AAs above the

586 cutoff threshold. Group 5 was consistently excluded as it only contains 2 highly divergent
587 sequences.

588 Comparison of Groups

589 Unique regions of each group of Aer homologs were identified by comparing the SHMR
590 scores for each pair-wise comparison with the overall conservation score of each AA and the
591 domain architecture of the Aer protein. Conservation scores were obtained from Jalview (59),
592 and along with the SHMR scores were smoothed and plotted using ggplot2 (62) in R. Smoothing
593 was performed by calculating the average of the 3 proceeding and following AAs for each
594 position. The domain architecture of Aer was obtained from the conserved domain database
595 using Aer from *P. aeruginosa* as a query (NP_250252.1) (CDD) (63). WebLogos were generated
596 using the Weblogo generator tool (64).

597 Distribution of Groups

598 Based on the ML tree grouping, the number of times a strain appeared in each group was
599 counted. This matrix was then transposed (now indicating the number of Aer homologs per
600 group for each strain). A hierarchically clustered heatmap using Bray-Curtis distance and
601 average clustering was made in R. Presence of Aer-2 (NP_248866.1) and CttP
602 (WP_003106690.1) were determined using BLAST searches specifically against the strains.
603 Pseudogenes were detected by BLASTing the nucleotide sequence of Aer.

604 Detection of Evidence of Horizontal Gene Transfer

605 Graphical representations of the Aer homologs nucleotide sequences were manually
606 inspected on NCBI. The first two genes upstream and downstream of the *aer* homolog were
607 noted, along with any mobile elements (transposase, integrase and inverted repeats) within 5kb.
608 For each different upstream and downstream gene their frequency of occurrence was calculated

609 for each homolog group. The frequency of occurrence of each type of mobile element was also
610 calculated, as well as for the complete set of homologs.

611 Generation of DNA Sequence Phylogeny

612 Gene sequences were obtained by BLAST (18) using sequences of *gyrB, rpoB* and *rpoD*
613 from *P. aeruginosa* PAO1. Sequences were aligned separately by MAFFT (65) and positions
614 with gaps in any strain were removed. Sequences from all three genes were then concatenated
615 and a ML phylogeny was generated using DIVEIN (66). DIVEIN parameters: substitution model,
616 GTR; equilibrium frequencies, optimized; proportion of invariable sites and gamma distribution
617 parameter, estimated; number of substitution rate categories, 4; tree searching, NNI+SPR; tree
618 optimized for topology and branch lengths; 100 bootstrap replicates. Tanglegrams were
619 generated in Dendroscope (67) by pruning branches from the genus DNA tree so that only strains
620 that matched the particular Aer homolog group were included.

621 Culture Growth

622 For molecular biology, cultures were routinely cultured in lysogeny broth (LB, 5 g/L
623 yeast extract, 10 g/L Tryptone, 10 g/L NaCl). For energy-taxis experiments *P. pseudoalcaligenes*
624 KF707 strains were grown overnight (16 h) in minimal salts media containing 10 mM pyruvate
625 or succinate. Minimal salts media contained (in g/L) K_2HPO_4, 3; NaH_2PO_4, 1.15; NH_4Cl, 1; KCl,
626 0.15; $MgSO_4$, 0.15; $CaCl_2$, 0.01; $FeSO_4$, 0.0025. The latter four were sterile filtered and added
627 after autoclaving.

628 Generation of Deletion Constructs and Mutants

629 Nucleotide sequences for Aer.g1, Aer.g2, Aer.g4, Aer2 and CttP were obtained from the
630 draft genome sequence of *P. pseudoalcaligenes* KF707 (68). Primers for a ~500bp region up and
631 downstream of each region were generated using Primer BLAST (69). Benchling was used to

alter the primers, adding BamHI or HinDIII restriction sites to the outer primers and the reverse-complement of the other inner primer to each of the internal primers. Genomic DNA was isolated by the phenol/chloroform method (70). In separate PCR reactions the upstream and downstream fragments were amplified using Hi-Fidelity (HF) Enzyme mix (Fisher Scientific, USA). Fragments were purified by gel extraction using an EZDNA kit (Omega Bio-Tek, USA) then pooled and used as the template for the second PCR reaction using only the outer primers. The pG19II vector(71), purified using an EZDNA plasmid mini kit II (Omega Bio-Tek, USA), and insert were digested using BamHI and HinDIII (Invitrogen, USA). Digestion products were purified then ligated together using T4 ligase (Invitrogen, USA. Ligations were transformed either directly into *E. coli* Top10F' or first into DH5α chemical competent cells using standard methods (70). White colonies were picked from LB X-Gal gentamycin (20 µg/mL) plates and their plasmids isolated and screened for a ~1kb bandshift. Those with the appropriate shift were sequenced (Eurofins, USA) to confirm the correct insert sequence. To delete the genes from the *P. pseudoalcaligenes* KF707 genome, the deletion construct containing plasmids were introduced by conjugation. Cultures of *E. coli* HB101 carrying the helper plasmid pRK2013 (72) and *E. coli* Top10F' carrying the deletion construct in pG19II were grown to early log-phase (OD ~0.3) along with the KF707 wild-type, or later, deletion mutants. Donor, helper and recipients were mixed and plated, grown overnight then the cell mass was collected and spread on AB glucose plates (5 g/L glucose, minimal salts media) containing 20 µg/mL gentamycin for 48h at 30°C.Transconjugant colonies were picked off the AB glucose Gm plates into LB no salt, LB 10% sucrose and LB + 20 µg/mL Gm. Colonies that were able to grow with Gm but not (at all) with sucrose were selected for continued use. The LB no salt overnight culture was used to inoculate LB 10% sucrose. After 4h, the culture was plated on LB 10% sucrose and colonies

were screened to determine if the deletion occurred. Colony PCR was used to find those that had ONLY a band at ~1Kbp which were sent for sequencing (Eurofins, USA) to confirm the deletion. *P. pseudoalcaligenes* KF707 *cheA::Tn5* from Tremaroli *et al* 2011 (29) is called the 'cheA' mutant throughout the text as its cheA1 (equivalent to *P. aeruginosa* PA01) has been disrupted.

Energy-taxis Swim Plates

Swim plates were made by making minimal salts media with 0.3% agar and 50mM succinate or pyruvate. Strains were grown overnight in minimal salts media containing 10mM of the appropriate carbon source. An inoculation needle was sterilized by ethanol and flaming then dipped into the overnight culture and carefully stabbed into the swim plate. The needle was re-inoculated for each stab and was re-sterilized for each strain. The diameter of growth for each strain was measured at the maximum distance away from the inoculation centre that bacteria were visible, at 24 and 48 h either manually using a ruler or digitally using a photograph and ImageJ (73). Experiments were repeated at least 3 times for all strains in each media, always including at least one wild-type to be used as a normalizing control. Collected data were processed in R to normalize the size of the growth diameter to the corresponding wild-type size at 24 and 48 h. The data were analyzed in a number of ways and it was determined that pooling the normalized data from 24 and 48 h to produce at least 6 replicates resulted in smaller error bars. Tukey's Honest Significant Differences test was used to determine if the differences between strains were significant for each carbon source. As the TukeyHSD() function in R (61) tests ALL pair-wise comparisons, only those comparing each strain to the wild-type are presented here. However, this makes these results more robust as the false positive correction for a confidence level of 0.95 was applied to all 325 comparisons which were tested.

678 Chemotaxis Swim Plates

679 Strains were grown up overnight as before, then 1 mL was pelleted, washed once with 1
680 mL minimal salts media (no carbon source) then resuspended in 100 µL minimal salts media (no
681 carbon source). 20 µL was spotted at the edge of a minimal salts media plate containing no
682 carbon source and 0.3% agar. Either an agar plug containing 50 mM carbon source or small
683 amount of crystals was placed in the centre and plates were incubated overnight at 30°C. Plates
684 were photographed and positive chemotaxis was interpreted as an arc of cells nearer to the centre
685 of the plate than the cells had been spotted.

686 Growth Curves

687 To ensure that the genetic manipulations had no effect on the growth of any of the strains,
688 growth was assayed over 24h. Overnight cultures of all strains were normalized to OD 0.1 then
689 diluted 1/100 into 200 µL minimal salts media with 10mM pyruvate in a 96 well microtiter plate.
690 The plate was incubated at 30°C, shaking at 150RPM and the OD600 was checked every 6h.

691 Complementation

692 Aer.g1 was complemented into the wild-type and the quintuple mutant (Aer-
693 2/CttP/Aer.g1/Aer.g2/Aer.g4) to demonstrate its importance in energy-taxis. Aer.g1 was cloned
694 by PCR amplifying the gene from genomic DNA using one of the deletion construct primers and
695 an additional primer that annealed ending within the start codon of Aer.g1. A HinDIII site was
696 added to this primer and the opposite outer primer from the deletion construct was used to
697 amplify the Aer.g1 gene. A pre-existing BamHI site 36nt from the end of the gene was taken
698 advantage of when the gel purified (EZDNA gel extraction kit, Omega Bio-Tek, USA) product
699 was digested alongside pSEVA_342 (74). Digested products were purified, ligated and
700 transformed into Top10F' the same as the deletion constructs. After confirming the insert by

sequencing (Eurofins, USA) the complementation and empty vector were transformed into the wild-type and quintuple deletion mutant as before, only 30µg/mL chloramphenicol was used to select for transformants. Resistant colonies of P. pseudoalcaligenes KF707 were picked into LB to make stocks. For the energy-taxis assays, overnight cultures were grown with 30µg/mL chloramphenicol. Instead of directly comparing to the wild-type, these strains were compared based on their change in energy-taxis diameters over 24h. All plasmids used in this study are summarized in Supplementary Table 6.

Acknowledgements

The authors would like to thank Stephana Cherak for her technical assistance, and Iain George and Dr. Gordon Chua for use of their imaging setup for photographing the swim plates. We thank Dr. Martina Cappelletti for assistance with the colony PCR design and her insights into KF707 physiology and genomic interpretation. This project was funded by Natural Sciences and Engineering Research Council (NSERC) of Canada Discovery Grant RGPIN/219895-2010 to RJT.

References

1. **Wadhams GH**, **Armitage JP**. 2004. Making sense of it all: Bacterial chemotaxis. Nat Rev Microbiol **5**:1024–1037.

2. **Sampedro I**, **Parales RE**, **Krell T**, **Hill JE**. 2015. Pseudomonas chemotaxis. FEMS Microbiol Rev **39**:17–46.

3. **Krell T**. 2015. Tackling the bottleneck in bacterial signal transduction research: high-throughput identification of signal molecules. Mol Microbiol **96**:685–8.

4. **Alexandre G**, **Greer-Phillips S**, **Zhulin IB**. 2004. Ecological role of energy taxis in microorganisms. FEMS Microbiol Rev **28**:113–26.

5. **Rebbapragada A**, **Johnson MS**, **Harding GP**, **Zuccarelli AJ**, **Fletcher HM**, **Zhulin IB**, **Taylor BL**. 1997. The Aer protein and the serine chemoreceptor Tsr independently sense intracellular energy levels and transduce oxygen, redox, and energy signals for Escherichia coli behavior. Proc Natl Acad Sci U S A **94**:10541–6.

6. **Bibikov SI**, **Biran R**, **Rudd KE**, **Parkinson JS**. 1997. A signal transducer for aerotaxis in Escherichia coli. J Bacteriol **179**:4075–4079.

7. **Samanta D**, **Widom J**, **Borbat PP**, **Freed JH**, **Crane BR**. 2016. Bacterial Energy Sensor Aer Modulates the Activity of the Chemotaxis Kinase CheA Based on the Redox State of the Flavin Cofactor. J Biol Chem **291**:25809–25814.

8. **Hong CS**, **Shitashiro M**, **Kuroda A**, **Ikeda T**, **Takiguchi N**, **Ohtake H**, **Kato J**. 2004. Chemotaxis proteins and transducers for aerotaxis in Pseudomonas aeruginosa. FEMS Microbiol Lett **231**:247–252.

9. **Sarand I**, **Österberg S**, **Holmqvist S**, **Holmfeldt P**, **Skärfstad E**, **Parales RE**, **Shingler V**. 2008. Metabolism-dependent taxis towards (methyl)phenols is coupled through the most abundant of three polar localized Aer-like proteins of Pseudomonas putida. Environ Microbiol **10**:1320–1334.

10. **Nichols NN**, **Harwood CS**. 2000. An aerotaxis transducer gene from Pseudomonas putida. FEMS Microbiol Lett **182**:177–183.

11. **Luu R a.**, **Schneider BJ**, **Ho CC**, **Nesteryuk V**, **Ngwesse SE**, **Liu X**, **Parales J V.**, **Ditty JL**, **Parales RE**. 2013. Taxis of Pseudomonas putida F1 toward phenylacetic acid is mediated by the energy taxis receptor AER2. Appl Environ Microbiol **79**:2416–2423.

12. **Watts KJ**, **Taylor BL**, **Johnson MS**. 2011. PAS/poly-HAMP signalling in Aer-2, a soluble haem-based sensor. Mol Microbiol **79**:686–99.

13. **Güvener ZT**, **Tifrea DF**, **Harwood CS**. 2006. Two different Pseudomonas aeruginosa chemosensory signal transduction complexes localize to cell poles and form and remould in stationary phase. Mol Microbiol **61**:106–18.

14. **Shitashiro M**, **Tanaka H**, **Hong CS**, **Kuroda A**, **Takiguchi N**, **Ohtake H**, **Kato J**. 2005. Identification of chemosensory proteins for trichloroethylene in Pseudomonas aeruginosa. J Biosci Bioeng **99**:396–402.

15. **Pham HT**, **Parkinson JS**. 2011. Phenol sensing by Escherichia coli chemoreceptors: a nonclassical mechanism. J Bacteriol **193**:6597–604.

16. **Bodilis J**, **Nsigue Meilo S**, **Cornelis P**, **De Vos P**, **Barray S**. 2011. A long-branch attraction artifact reveals an adaptive radiation in pseudomonas. Mol Biol Evol **28**:2723–6.

17. **Gomila M**, **Peña A**, **Mulet M**, **Lalucat J**, **García-Valdés E**. 2015. Phylogenomics and systematics in Pseudomonas. Front Microbiol **6**:214.

18. **Altschul SF**, **Gish W**, **Miller W**, **Myers EW**, **Lipman DJ**. 1990. Basic local alignment search tool. J Mol Biol **215**:403–10.

19. **Papadopoulos JS**, **Agarwala R**. 2007. COBALT: Constraint-based alignment tool for multiple protein sequences. Bioinformatics **23**:1073–1079.

20. **Guindon S**, **Gascuel O**. 2003. A Simple, Fast, and Accurate Algorithm to Estimate Large Phylogenies by Maximum Likelihood. Syst Biol **52**:696–704.

21. **Guindon S**, **Dufayard J**, **Lefort V**. 2010. New Algorithms and Methods to Estimate Maximum-Likelihood Phylogenies : Assessing the Performance of PhyML 3 . 0. Syst Biol **59**:307–21.

22. **Brandt BW**, **Feenstra KA**, **Heringa J**. 2010. Multi-Harmony: detecting functional specificity from sequence alignment. Nucleic Acids Res **38**:W35-40.

23. **Özen AI**, **Ussery DW**. 2012. Defining the Pseudomonas genus: where do we draw the line with Azotobacter? Microb Ecol **63**:239–48.

24. **Schultz J**, **Copley RR**, **Doerks T**, **Ponting CP**, **Bork P**. 2000. SMART: a web-based tool for the study of genetically mobile domains. Nucleic Acids Res **28**:231–4.

25. **Garcia D**, **Watts KJ**, **Johnson MS**, **Taylor BL**. 2016. Delineating PAS-HAMP interaction surfaces and signalling-associated changes in the aerotaxis receptor Aer. Mol Microbiol **100**:156–172.

780 26. **Alexander RP**, **Zhulin IB**. 2007. Evolutionary genomics reveals conserved structural
781 determinants of signaling and adaptation in microbial chemoreceptors. Proc Natl Acad Sci
782 U S A **104**:2885–90.

783 27. **Hmelo LR**, **Borlee BR**, **Almblad H**, **Love ME**, **Randall TE**, **Tseng BS**, **Lin C**, **Irie Y**,
784 **Storek KM**, **Yang JJ**, **Siehnel RJ**, **Howell PL**, **Singh PK**, **Tolker-Nielsen T**, **Parsek**
785 **MR**, **Schweizer HP**, **Harrison JJ**. 2015. Precision-engineering the Pseudomonas
786 aeruginosa genome with two-step allelic exchange. Nat Protoc **10**:1820–1841.

787 28. **Harwood CS**, **Nichols NN**, **Kim M**, **Diitty JL**, **Parales RE**. 1994. Identification of the
788 pcaRKF Gene Cluster from Pseudomonas. Biotechniques **176**:6479–6488.

789 29. **Tremaroli V**, **Fedi S**, **Tamburini S**, **Viti C**, **Tatti E**, **Ceri H**, **Turner RJ**, **Zannoni D**.
790 2011. A histidine-kinase cheA gene of Pseudomonas pseudoalcaligens KF707 not only
791 has a key role in chemotaxis but also affects biofilm formation and cell metabolism.
792 Biofouling **27**:33–46.

793 30. **Nichols NN**, **Harwood CS**. 2000. An aerotaxis transducer gene from Pseudomonas putida.
794 FEMS Microbiol Lett **182**:177–183.

795 31. **Sarand I**, **Österberg S**, **Holmqvist S**, **Holmfeldt P**, **Skärfstad E**, **Parales RE**, **Shingler**
796 **V**. 2008. Metabolism-dependent taxis towards (methyl)phenols is coupled through the
797 most abundant of three polar localized Aer-like proteins of Pseudomonas putida. Environ
798 Microbiol **10**:1320–1334.

799 32. **Luu R a.**, **Schneider BJ**, **Ho CC**, **Nesteryuk V**, **Ngwesse SE**, **Liu X**, **Parales J V.**, **Ditty**
800 **JL**, **Parales RE**. 2013. Taxis of Pseudomonas putida F1 toward phenylacetic acid is

mediated by the energy taxis receptor AER2. Appl Environ Microbiol **79**:2416–2423.

33. **García-Fontana C**, **Corral Lugo A**, **Krell T**. 2014. Specificity of the CheR2 methyltransferase in Pseudomonas aeruginosa is directed by a C-terminal pentapeptide in the McpB chemoreceptor. Sci Signal **7**:ra34.

34. **Henry JT**, **Crosson S**. 2011. Ligand-Binding PAS Domains in a Genomic, Cellular, and Structural Context. Annu Rev Microbiol **65**:261–286.

35. **Ortega DR**, **Fleetwood AD**, **Krell T**, **Harwood CS**, **Jensen GJ**, **Zhulin IB**. 2017. Assigning chemoreceptors to chemosensory pathways in Pseudomonas aeruginosa. Proc Natl Acad Sci 201708842.

36. **Taylor BL**. 2007. Aer on the inside looking out: paradigm for a PAS?HAMP role in sensing oxygen, redox and energy. Mol Microbiol **65**:1415–1424.

37. **Parkinson JS**. 2010. Signaling mechanisms of HAMP domains in chemoreceptors and sensor kinases. Annu Rev Microbiol **64**:101–122.

38. **Williams KP**, **Gillespie JJ**, **Sobral BWS**, **Nordberg EK**, **Snyder EE**, **Shallom JM**, **Dickerman AW**. 2010. Phylogeny of gammaproteobacteria. J Bacteriol **192**:2305–14.

39. **Behrens W**, **Schweinitzer T**, **McMurry JL**, **Loewen PC**, **Buettner FFR**, **Menz S**, **Josenhans C**. 2016. Localisation and protein-protein interactions of the Helicobacter pylori taxis sensor TlpD and their connection to metabolic functions. Sci Rep **6**:23582.

40. **Pei J**, **Mitchell DA**, **Dixon JE**, **Grishin N V.** 2011. Expansion of Type II CAAX Proteases Reveals Evolutionary Origin of γ-Secretase Subunit APH-1Journal of Molecular Biology.

822 41. **Nishiyama S**, **Ohno S**, **Ohta N**, **Inoue Y**, **Fukuoka H**, **Ishijima A**, **Kawagishi I**. 2010.
823 Thermosensing function of the Escherichia coli redox sensor Aer. J Bacteriol **192**:1740–3.

824 42. **Yang Y**, **Sourjik V**. 2012. Opposite responses by different chemoreceptors set a tunable
825 preference point in *Escherichia coli* pH taxis. Mol Microbiol **86**:1482–1489.

826 43. **Bardy SL**, **Briegel A**, **Rainville S**, **Krell T**. 2017. Recent advances and future prospects
827 in bacterial and archaeal locomotion and signal transduction. J Bacteriol **199**:e00203-17.

828 44. **Österberg S**, **Åberg A**, **Herrera Seitz MK**, **Wolf-Watz M**, **Shingler V**. 2013. Genetic
829 dissection of a motility-associated c-di-GMP signalling protein of Pseudomonas putida.
830 Environ Microbiol Rep **5**:556–65.

831 45. **Xu L**, **Xin L**, **Zeng Y**, **Yam JKH**, **Ding Y**, **Venkataramani P**, **Cheang QW**, **Yang X**,
832 **Tang X**, **Zhang L-H**, **Chiam K-H**, **Yang L**, **Liang Z-X**. 2016. A cyclic di-GMP–binding
833 adaptor protein interacts with a chemotaxis methyltransferase to control flagellar motor
834 switching. Sci Signal **9**.

835 46. **Zhulin IB**. 2001. The superfamily of chemotaxis transducers: From physiology to
836 genomics and back. Adv Microb Physiol **45**:157–198.

837 47. **Borziak K**, **Fleetwood AD**, **Zhulin IB**. 2013. Chemoreceptor gene loss and acquisition
838 via horizontal gene transfer in Escherichia coli. J Bacteriol **195**:3596–602.

839 48. **Qiu X**, **Kulasekara BR**, **Lory S**. 2009. Role of Horizontal Gene Transfer in the Evolution
840 of Pseudomonas aeruginosa Virulence. Genome Dyn **6**:126–39.

841 49. **Sengeløv G**, **Kristensen KJ**, **Sørensen AH**, **Kroer N**, **Sørensen SJ**. 2001. Effect of
842 Genomic Location on Horizontal Transfer of a Recombinant Gene Cassette Between

Pseudomonas Strains in the Rhizosphere and Spermosphere of Barley Seedlings. Curr Microbiol **42**:160–167.

50. **Tremaroli V**, **Suzzi CV**, **Fedi S**, **Ceri H**, **Zannoni D**, **Turner RJ**. 2010. Tolerance of Pseudomonas pseudoalcaligenes KF707 to metals, polychlorobiphenyls and chlorobenzoates: Effects on chemotaxis-, biofilm- and planktonic-grown cells. FEMS Microbiol Ecol **74**:291–301.

51. **Fedi S**, **Triscari-Barberi T**, **Nappi MR**, **Sandri F**, **Booth SC**, **Turner RJ**, **Attimonelli M**, **Cappelletti M**, **Zannoni D**. 2016. The Role of cheA Genes in Swarming and Swimming Motility ofPseudomonas pseudoalcaligenes KF707. Microbes Environ **31**:169–172.

52. **Yao J**, **Allen C**. 2007. The plant pathogen Ralstonia solanacearum needs aerotaxis for normal biofilm formation and interactions with its tomato host. J Bacteriol **189**:6415–6424.

53. **Boin MA**, **Häse CC**. 2007. Characterization of Vibrio cholerae aerotaxis. FEMS Microbiol Lett **276**:193–201.

54. **Ferrandez A**, **Hawkins AC**, **Summerfield DT**, **Harwood CS**. 2002. Cluster II che Genes from Pseudomonas aeruginosa Are Required for an Optimal Chemotactic Response. J Bacteriol **184**:4374–4383.

55. **Kim HE**, **Shitashiro M**, **Kuroda A**, **Takiguchi N**, **Ohtake H**, **Kato J**. 2006. Identification and characterization of the chemotactic transducer in Pseudomonas aeruginosa PAO1 for positive chemotaxis to trichloroethylene. J Bacteriol **188**:6700–6702.

56. **Ortega Á**, **Zhulin IB**, **Krell T**. 2017. Sensory Repertoire of Bacterial Chemoreceptors.

Microbiol Mol Biol Rev **81**:e00033-17.

57. **Tatusova T**, **Ciufo S**, **Fedorov B**, **O'Neill K**, **Tolstoy I**. 2014. RefSeq microbial genomes database: new representation and annotation strategy. Nucleic Acids Res **42**:D553-9.

58. **Cochrane G**, **Karsch-Mizrachi I**, **Takagi T**, **International Nucleotide Sequence Database Collaboration IN**. 2016. The International Nucleotide Sequence Database Collaboration. Nucleic Acids Res **44**:D48-50.

59. **Waterhouse AM**, **Procter JB**, **Martin DMA**, **Clamp M**, **Barton GJ**. 2009. Jalview Version 2--a multiple sequence alignment editor and analysis workbench. Bioinformatics **25**:1189–91.

60. **Edgar RC**. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res **32**:1792–1797.

61. **R Core Team**. 2014. R: A language and environment for statistical computing. R Found Stat Comput Vienna, Austria 2014.

62. **Wickham H**. 2011. ggplot2. Wiley Interdiscip Rev Comput Stat **3**:180–185.

63. **Marchler-Bauer A**, **Derbyshire MK**, **Gonzales NR**, **Lu S**, **Chitsaz F**, **Geer LY**, **Geer RC**, **He J**, **Gwadz M**, **Hurwitz DI**, **Lanczycki CJ**, **Lu F**, **Marchler GH**, **Song JS**, **Thanki N**, **Wang Z**, **Yamashita RA**, **Zhang D**, **Zheng C**, **Bryant SH**. 2015. CDD: NCBI's conserved domain database. Nucleic Acids Res **43**:D222-6.

64. **Crooks GE**, **Hon G**, **Chandonia JM**, **Brenner SE**. 2004. WebLogo: A sequence logo generator. Genome Res **14**:1188–1190.

884 65. **Katoh K**, **Standley DM**. 2013. MAFFT Multiple Sequence Alignment Software Version
885 7: Improvements in Performance and Usability. Mol Biol Evol **30**:772–780.

886 66. **Deng W**, **Maust BS**, **Nickle DC**, **Learn GH**, **Liu Y**, **Heath L**, **Pond SLK**, **Mullins JI**.
887 2010. DIVEIN: A web server to analyze phylogenies, sequence divergence, diversity, and
888 informative sites. Biotechniques **48**:405–408.

889 67. **Huson DH**, **Scornavacca C**. 2012. Dendroscope 3: An Interactive Tool for Rooted
890 Phylogenetic Trees and Networks. Syst Biol **61**:1061–1067.

891 68. **Triscari-Barberi T**, **Simone D**, **Calabrese FM**, **Attimonelli M**, **Hahn KR**, **Amoako KK**,
892 **Turner RJ**, **Fedi S**, **Zannoni D**. 2012. Genome sequence of the polychlorinated-biphenyl
893 degrader pseudomonas pseudoalcaligenes KF707. J Bacteriol **194**:4426–4427.

894 69. **Ye J**, **Coulouris G**, **Zaretskaya I**, **Cutcutache I**, **Rozen S**, **Madden TL**. 2012. Primer-
895 BLAST: a tool to design target-specific primers for polymerase chain reaction. BMC
896 Bioinformatics **13**:134.

897 70. **Sambrook J**, **Fritsch EF**, **Maniatis T**. 1989. No Title. Mol Biol A Lab Man.

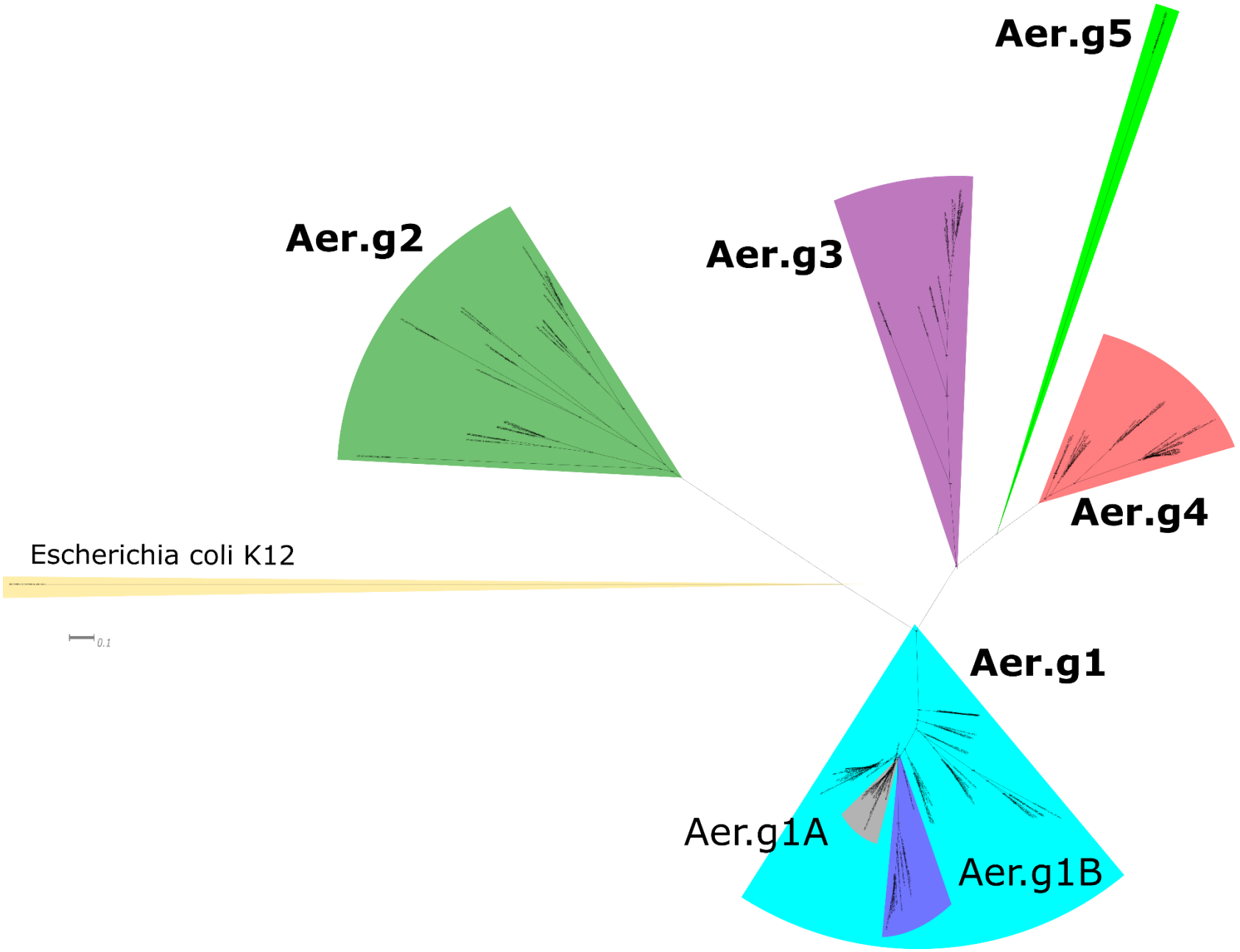
898 71. **Maseda H**, **Sawada I**, **Saito K**, **Uchiyama H**, **Nakae T**, **Nomura N**. 2004. Enhancement
899 of the mexAB-oprM efflux pump expression by a quorum-sensing autoinducer and its
900 cancellation by a regulator, MexT, of the mexEF-oprN efflux pump operon in
901 Pseudomonas aeruginosa. Antimicrob Agents Chemother **48**:1320–8.

902 72. **Figurski DH**, **Helinski DR**. 1979. Replication of an origin-containing derivative of
903 plasmid RK2 dependent on a plasmid function provided in trans. Proc Natl Acad Sci U S
904 A **76**:1648–52.

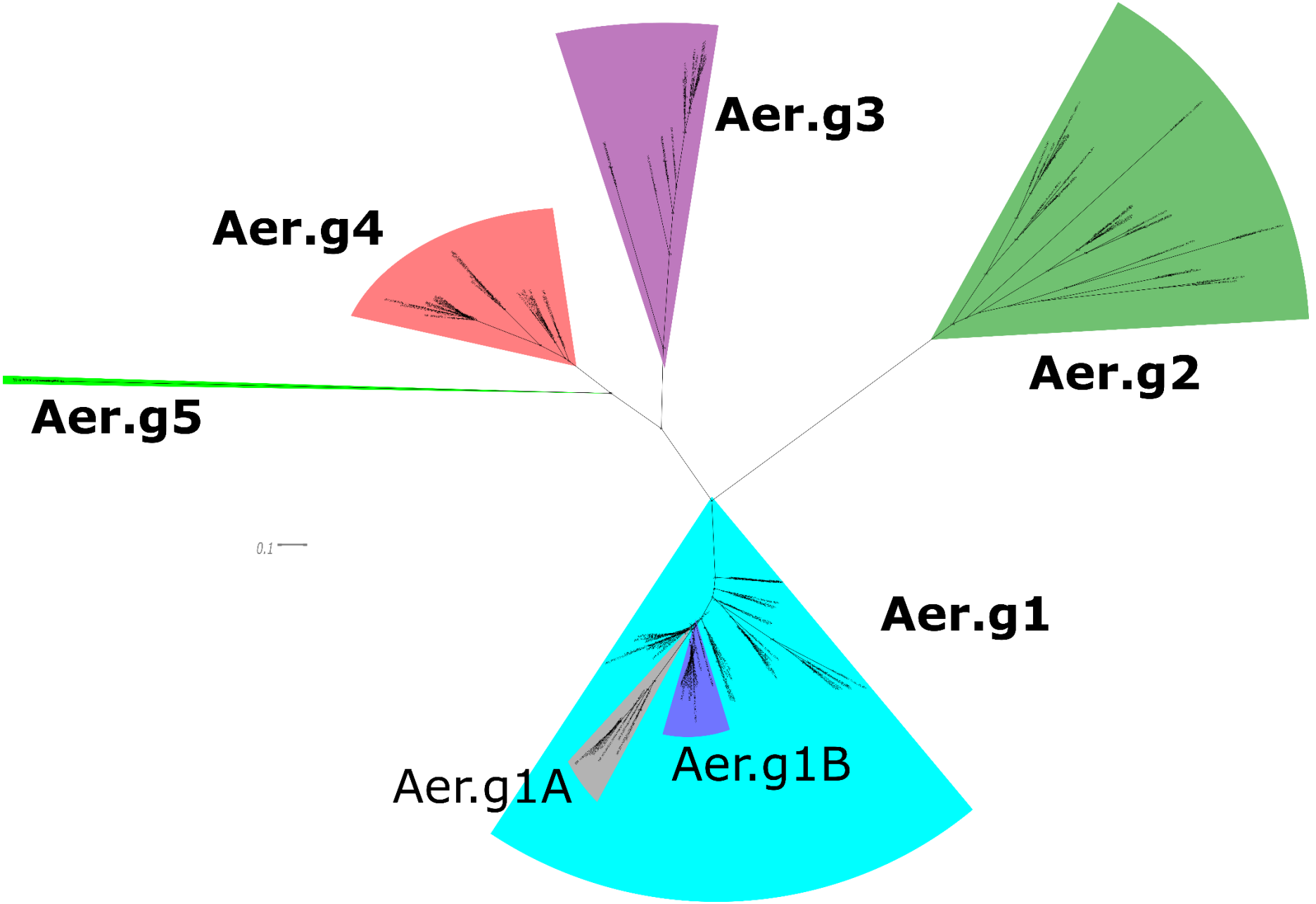
905 73. **Schneider CA**, **Rasband WS**, **Eliceiri KW**. 2012. NIH Image to ImageJ: 25 years of
906 image analysis. Nat Methods **9**:671–675.

907 74. **Silva-Rocha R**, **Martínez-García E**, **Calles B**, **Chavarría M**, **Arce-Rodríguez A**, **De**
908 **Las Heras A**, **Páez-Espino AD**, **Durante-Rodríguez G**, **Kim J**, **Nikel PI**, **Platero R**, **De**
909 **Lorenzo V**. 2013. The Standard European Vector Architecture (SEVA): A coherent
910 platform for the analysis and deployment of complex prokaryotic phenotypes. Nucleic
911 Acids Res **41**:D666–D675.

912

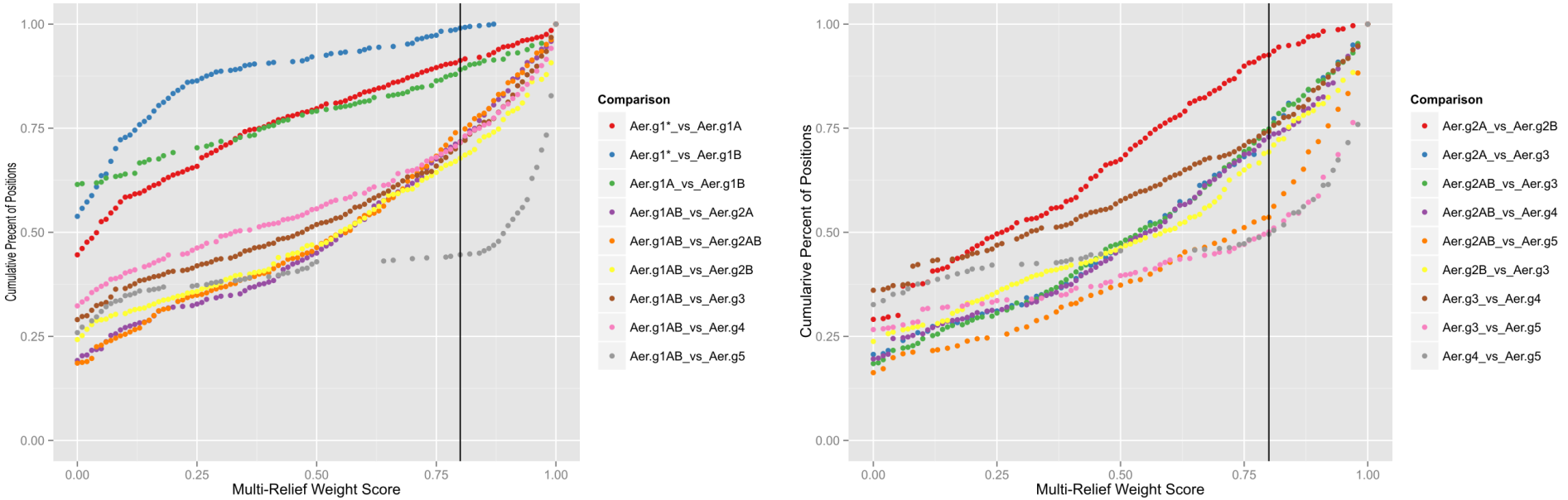


Supplementary Figure 1: Maximum-Likelihood consensus tree showing phylogenetic relationship between Aer protein sequences from *Pseudomonas* species, including Aer sequence from *E. coli*. Sequences were grouped according to branching pattern and inspection of the alignment, and were confirmed in subsequent analysis, see following figures and text for details. Numbers along branches indicate bootstrap support values from 100 replicates. Scale bar indicates branch length equivalent to 0.1 AA substitutions per site.

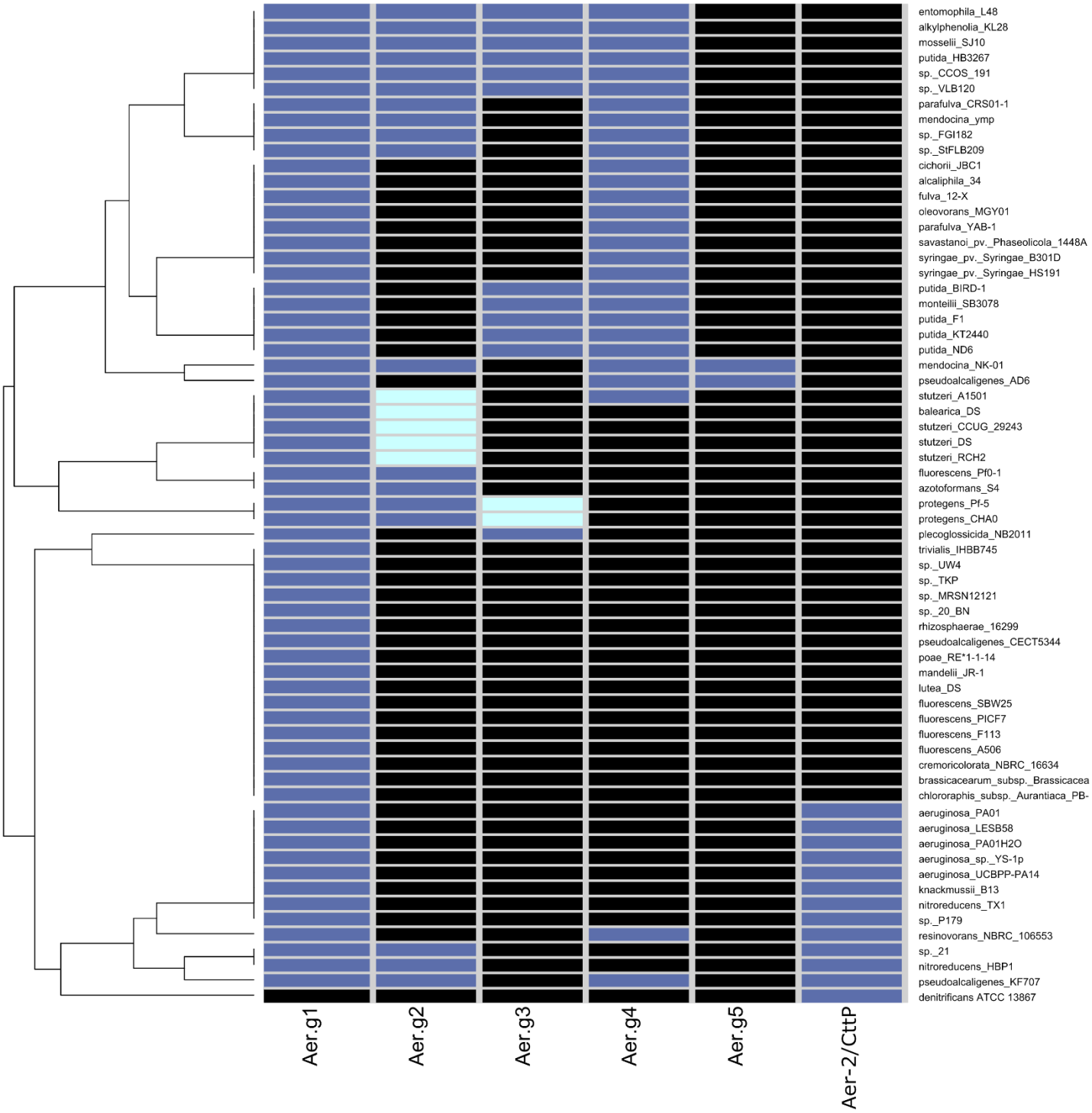


Supplementary Figure 2: Maximum-Likelihood consensus tree showing phylogenetic relationship between Aer protein sequences from *Pseudomonas* species. Sequences were grouped according to branching pattern and inspection of the alignment, and were confirmed in subsequent analysis, see following figures and text for details. Numbers along branches indicate bootstrap support values from 100 replicates. Scale bar indicates branch length equivalent to 0.1 AA substitutions per site.

Supplementary Figure 3: Maximum-Likelihood consensus tree showing phylogenetic relationship between Aer protein sequences from *Pseudomonas* species. Sequences were grouped according to branching pattern and inspection of the alignment, and were confirmed in subsequent analysis, see following figures and text for details. The same characterized Aer sequences (red stars), sequences from *P. pseudoalcaligenes* KF707 (black stars) as Figure 1 are marked here. This tree is equivalent to the cladogram in Figure 1, only branch lengths and all labels are maintained. Tree was generated unrooted, then the root placed based on comparison between the unrooted tree and a tree rooted to Aer from *E. coli*, see Supplementary Figures 1 and 2 for details. Numbers along branches indicate bootstrap support values from 100 replicates. Scale bar indicates branch length equivalent to 0.1 AA substitutions per site.

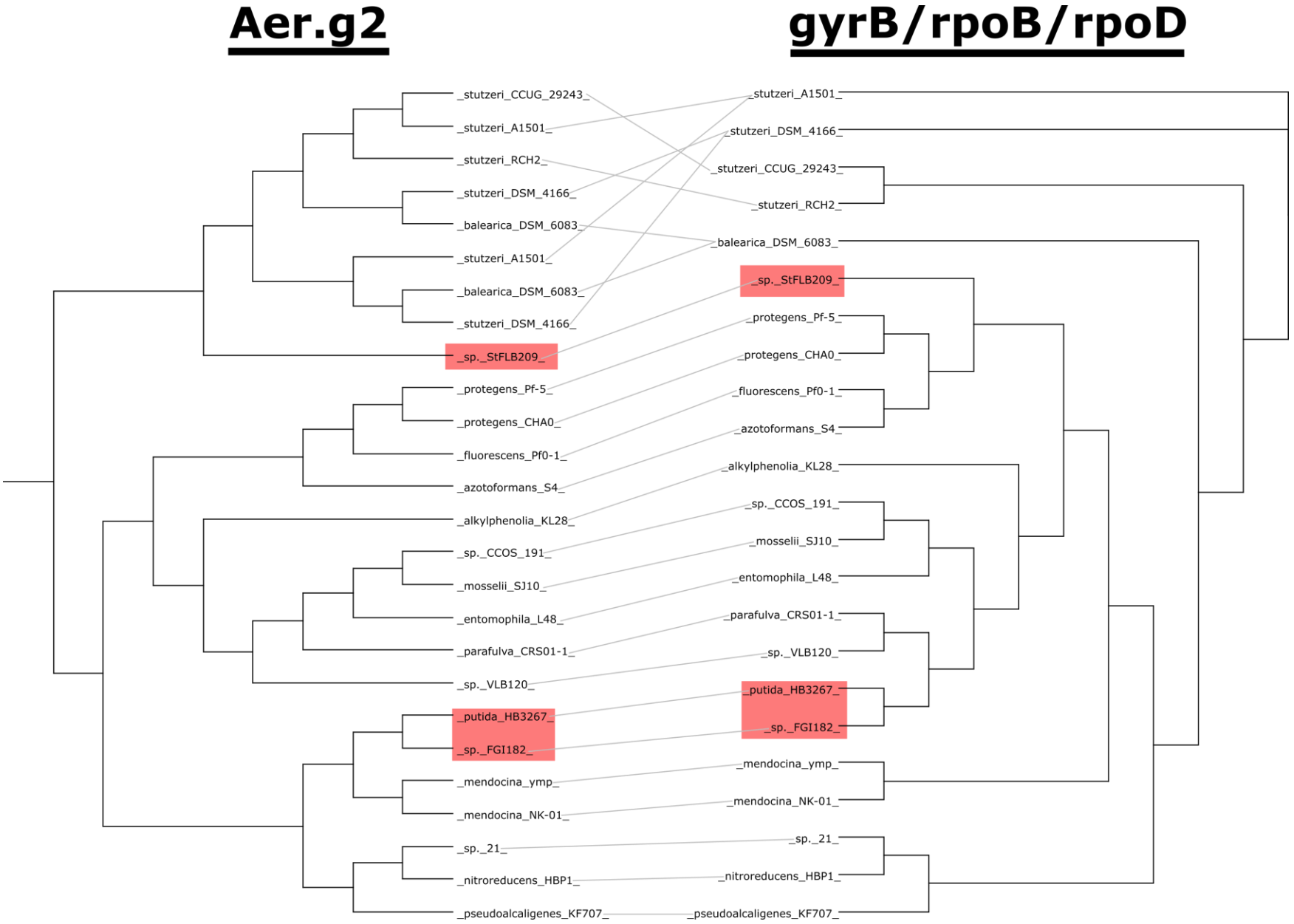


Supplementary Figure 4: Empirical cumulative distribution functions showing percent of amino acid positions at each multi-relief score value for comparisons of Aer homolog groups. Multi-relief weight scores ≥ 0.8 (black vertical line) indicate positions that are conserved within groups but not between groups. The point on the y-axis where each group comparison intersects with this line indicates the percentage of positions below the cutoff. Most group definitions result in ~>25% unique positions. Comparisons with group 5 show >50% unique positions (Aer.g5 only has two sequences). Intragroup comparisons between 1A/1B and 2A/2B show that these comparisons only result in <15% unique positions meaning their inclusion in their mother group is justified. Aer.g1* indicates Aer.g1 sequences not including Aer.g1A and Aer.g1B. Aer.g1AB includes all Aer.g1 sequences including Aer.g1A and Aer.g1B.

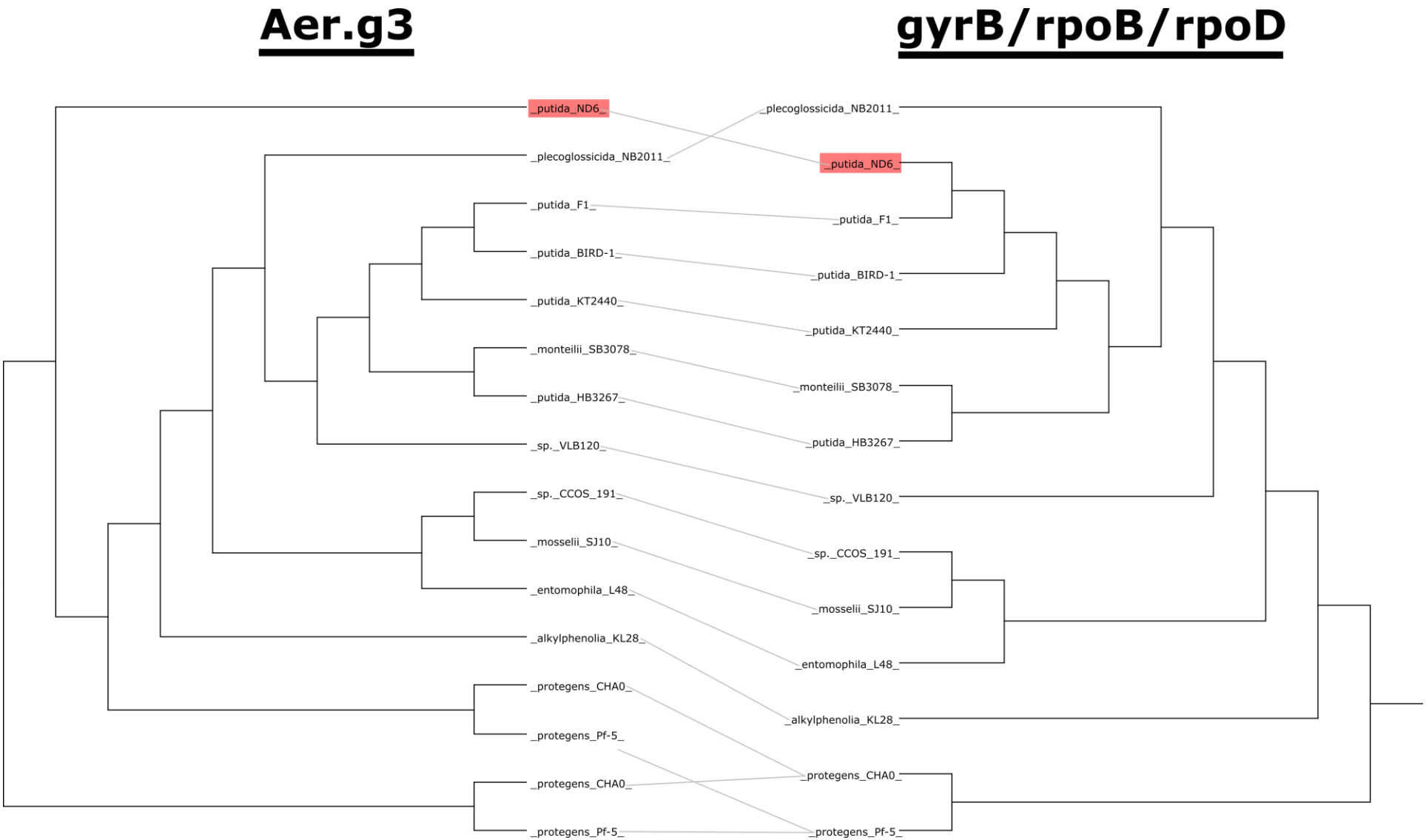


Supplementary Figure 5: Hierarchically clustered heatmap showing the presence and number of Aer homologs, Aer-2 and Cttp in select *Pseudomonas* species. Species were clustered using the Bray-Curtis distance metric and average linkages. The number of homologs that each strain possess is indicated by the box colour: zero (black), one (blue), two (cyan).

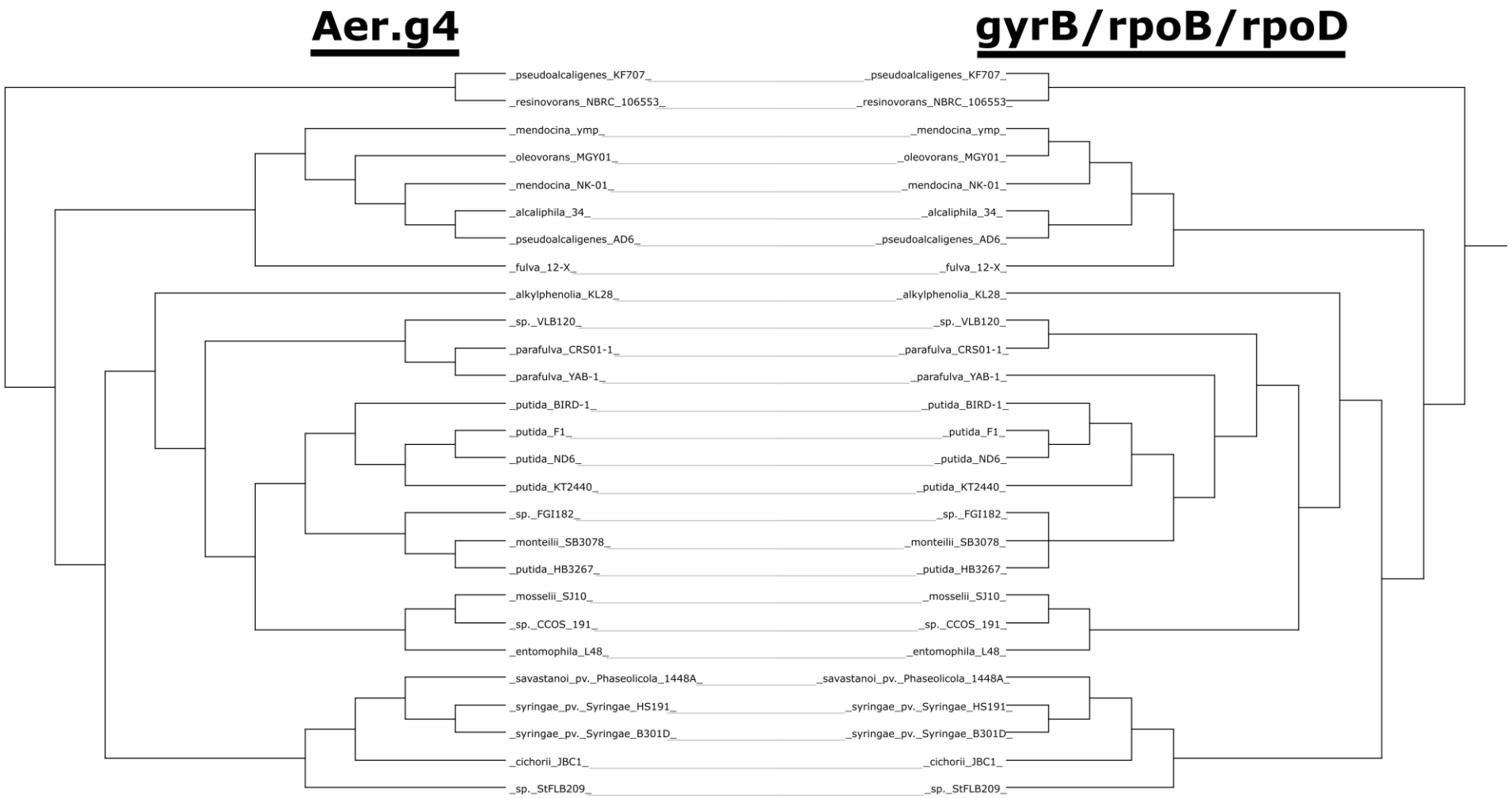
Supplementary Figure 6: Tanglegram showing tree matching between Aer.g1 subgroup tree and *Pseudomonas* genus phylogeny. Only species with a matching Aer sequence are included in the genus phylogeny. Matches highlighted in red indicate incongruity between the *gyrB/rpoB/rpoD* nucleotide sequence based tree and the Aer protein sequence based tree.



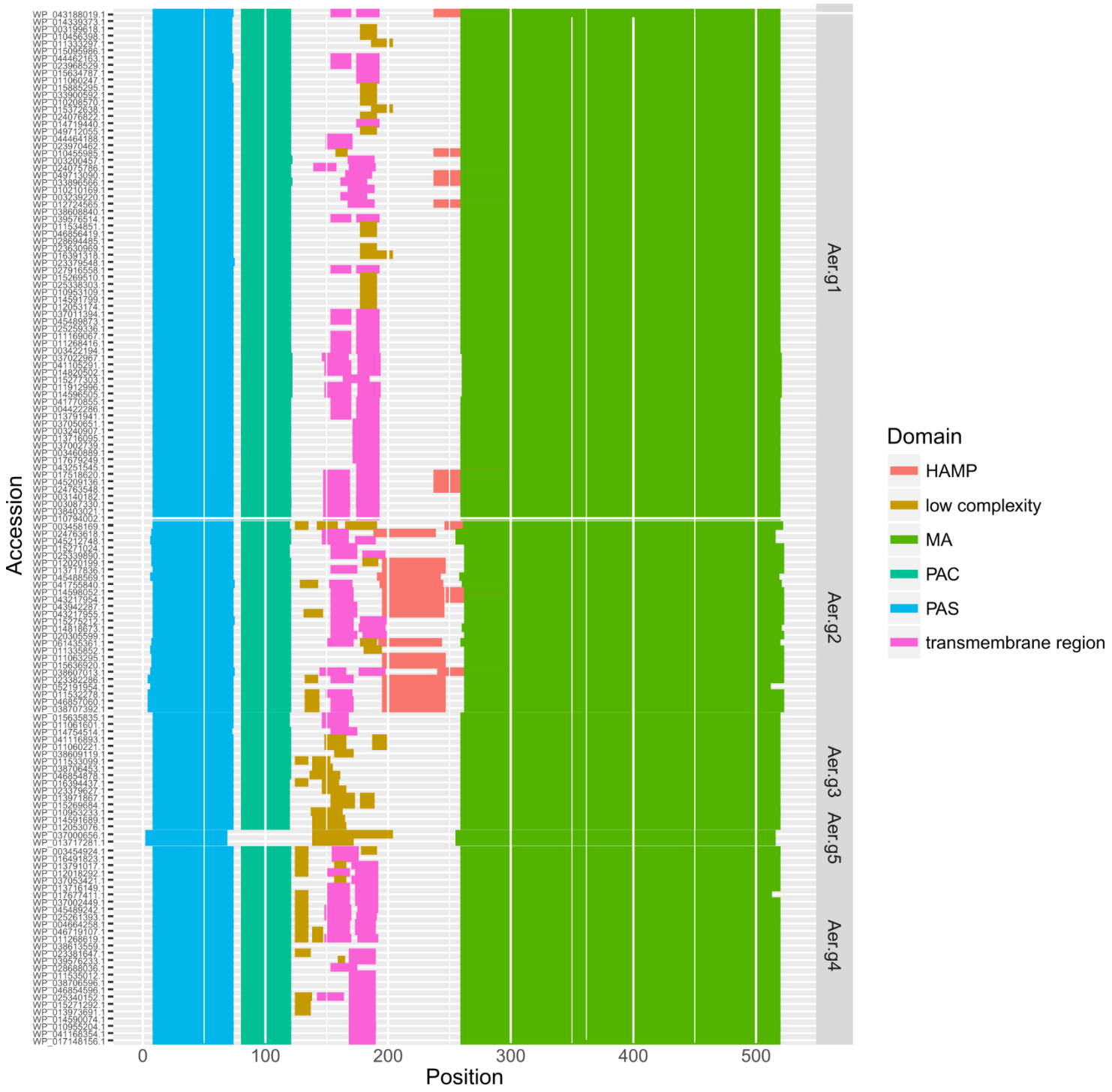
Supplementary Figure 7: Tanglegram showing tree matching between Aer.g2 subgroup tree and *Pseudomonas* genus phylogeny. Only species with a matching Aer sequence are included in the genus phylogeny. Matches highlighted in red indicate incongruity between the *gyrB/rpoB/rpoD* nucleotide sequence based tree and the Aer protein sequence based tree.



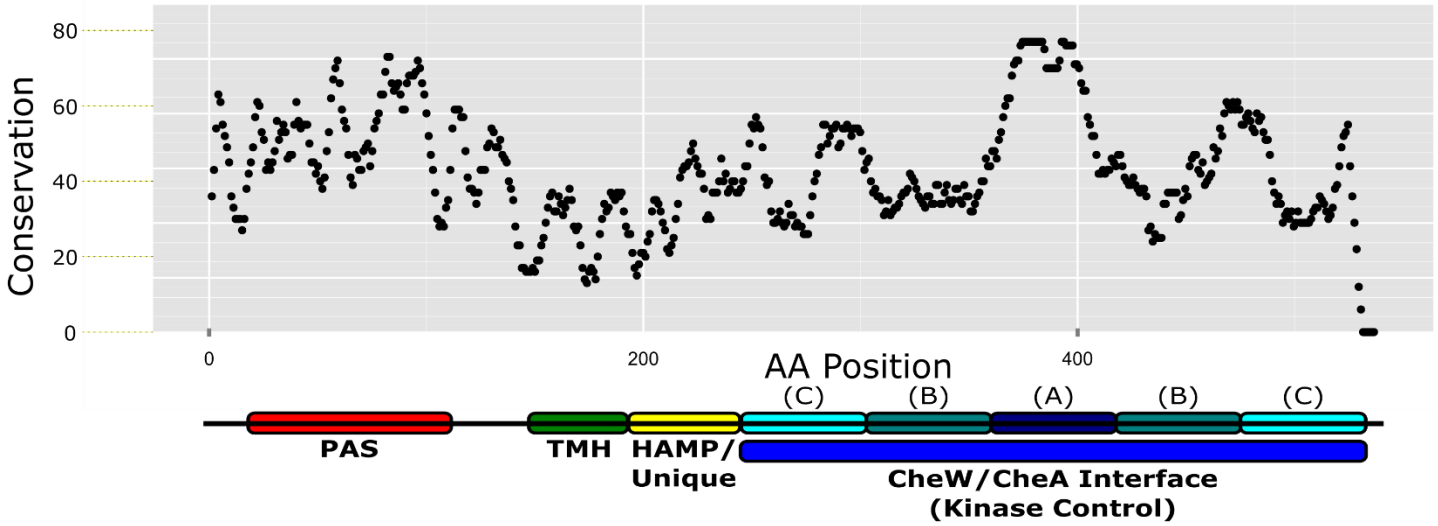
Supplementary Figure 8: Tanglegram showing tree matching between Aer.g3 subgroup tree and *Pseudomonas* genus phylogeny. Only species with a matching Aer sequence are included in the genus phylogeny. Matches highlighted in red indicate incongruity between the *gyrB/rpoB/rpoD* nucleotide sequence based tree and the Aer protein sequence based tree.



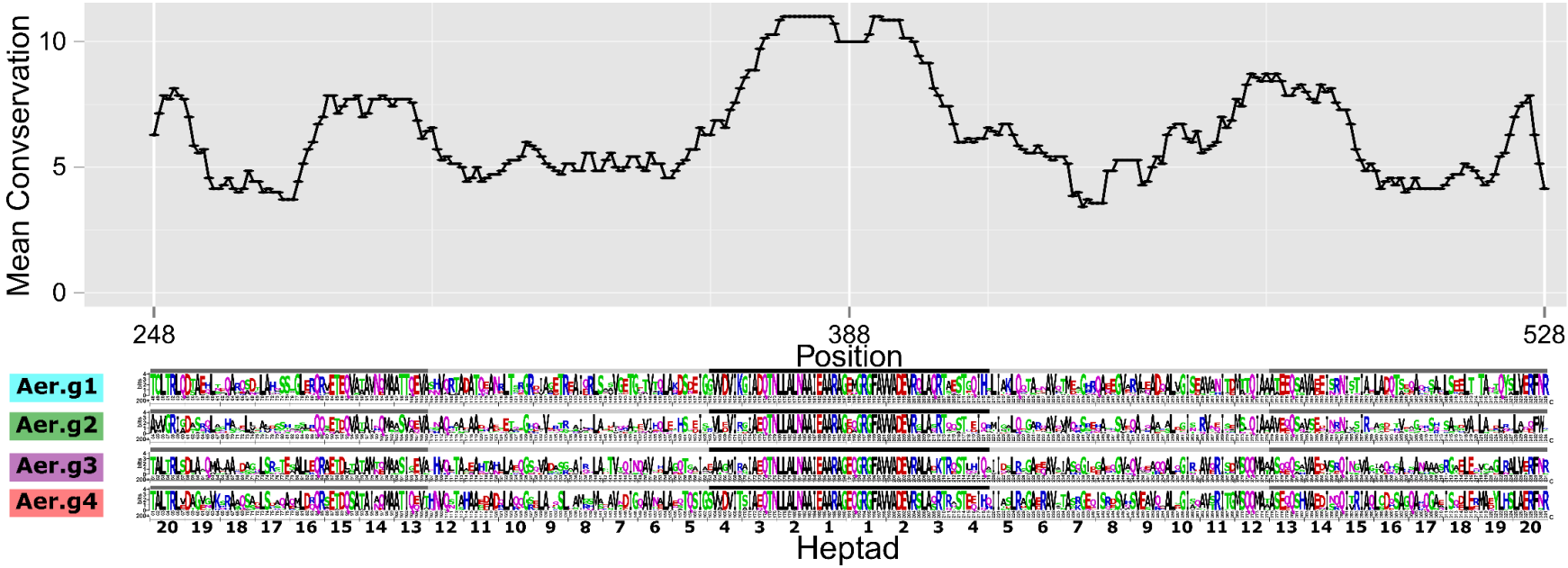
Supplementary Figure 9: Tanglegram showing tree matching between Aer.g1 subgroup tree and *Pseudomonas* genus phylogeny. Only species with a matching Aer sequence are included in the genus phylogeny. Matches highlighted in red indicate incongruity between the *gyrB/rpoB/rpoD* nucleotide sequence based tree and the Aer protein sequence based tree.



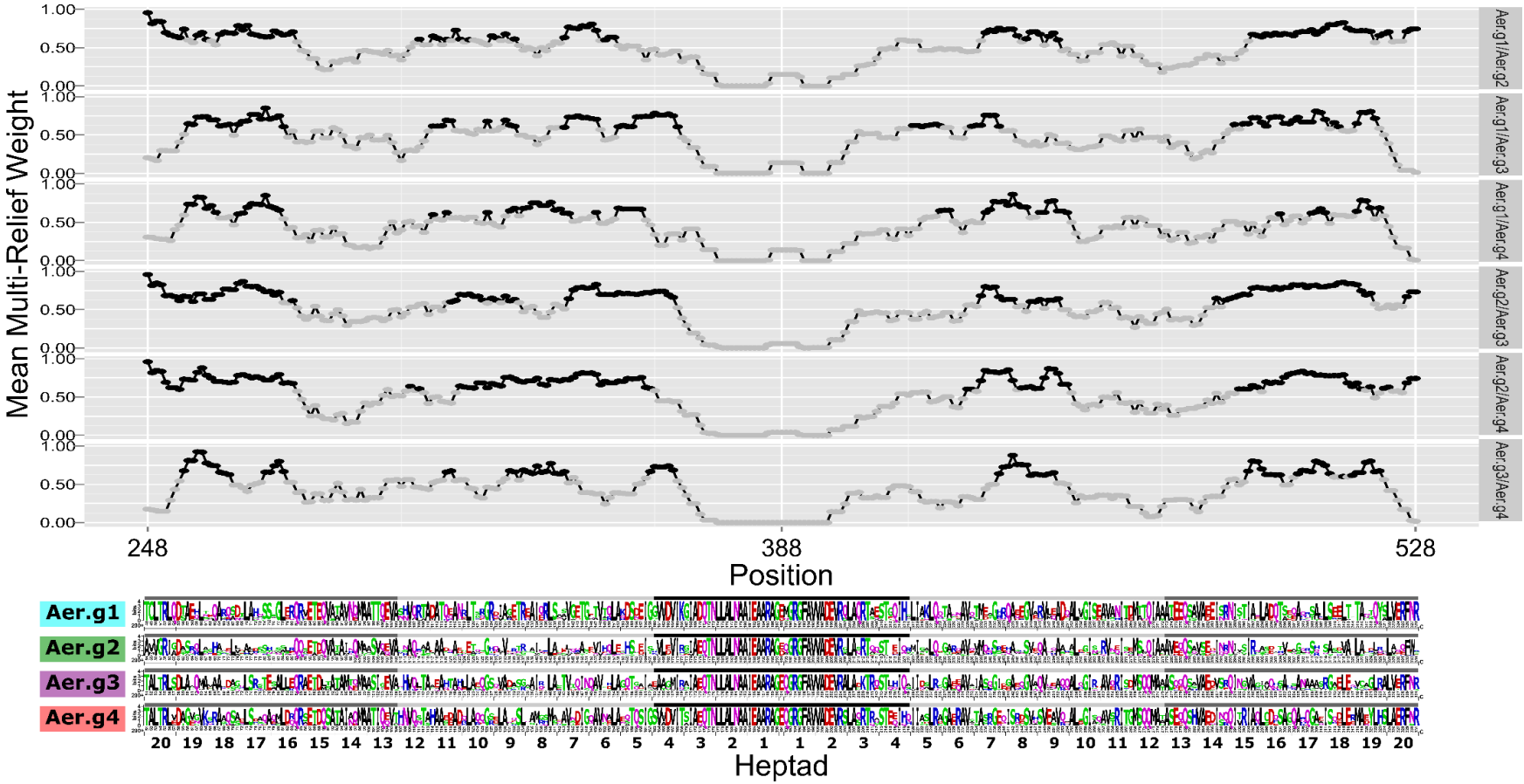
Supplementary Figure 10: Domain architecture of Aer sequences from select *Pseudomonas* species. Sequences were submitted to the SMART database and the results collected (1). Domain start and end positions were used to mark coloured bars: PAS (Pern/Art/Sin, blue); PAC (PAS associated domain, teal); transmembrane region (pink); HAMP (Histidine kinase/adenylyl cyclases/methyl-accepting chemotaxis proteins/phosphatases, red); MA (methyl-accepting, green); low-complexity (brown). The methyl-accepting domain is also called the kinase control module or CheW/CheA interface.



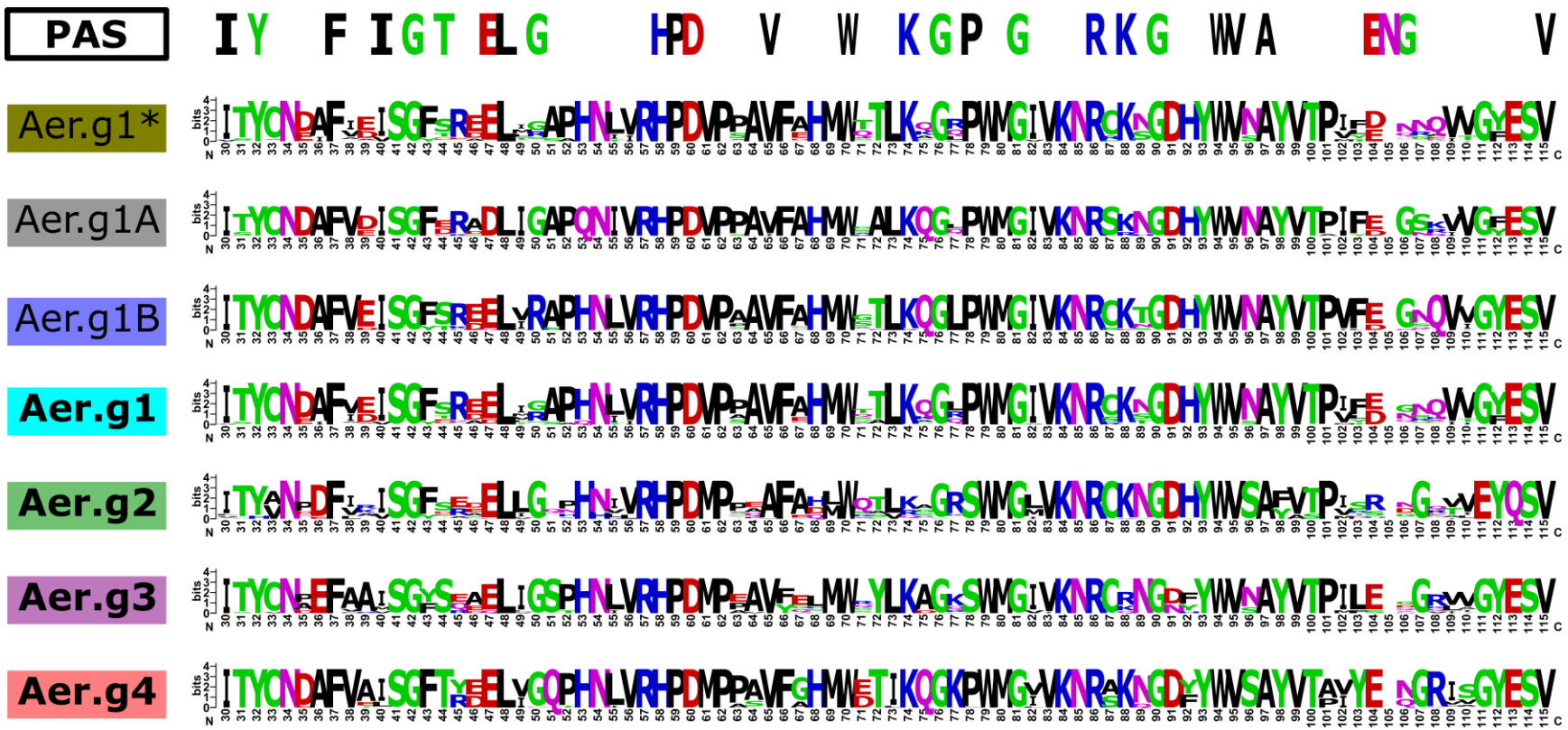
Supplementary Figure 11: Aer domain architecture and conservation across the entire length of all Aer sequences. Values were smoothed by taking the average of a position and the 3 preceding and following positions. Domains are PAS (Pern/Art/Sin) ligand binding (red), TMH (transmembrane helices, green), HAMP/Unique (in Aer.g2 HAMP, unique domains in other groups, yellow), CheW/CheA Interface also called Kinase Control (subdivided into A, signaling, dark blue; B, flexible bundle, teal; C, methylation, cyan).



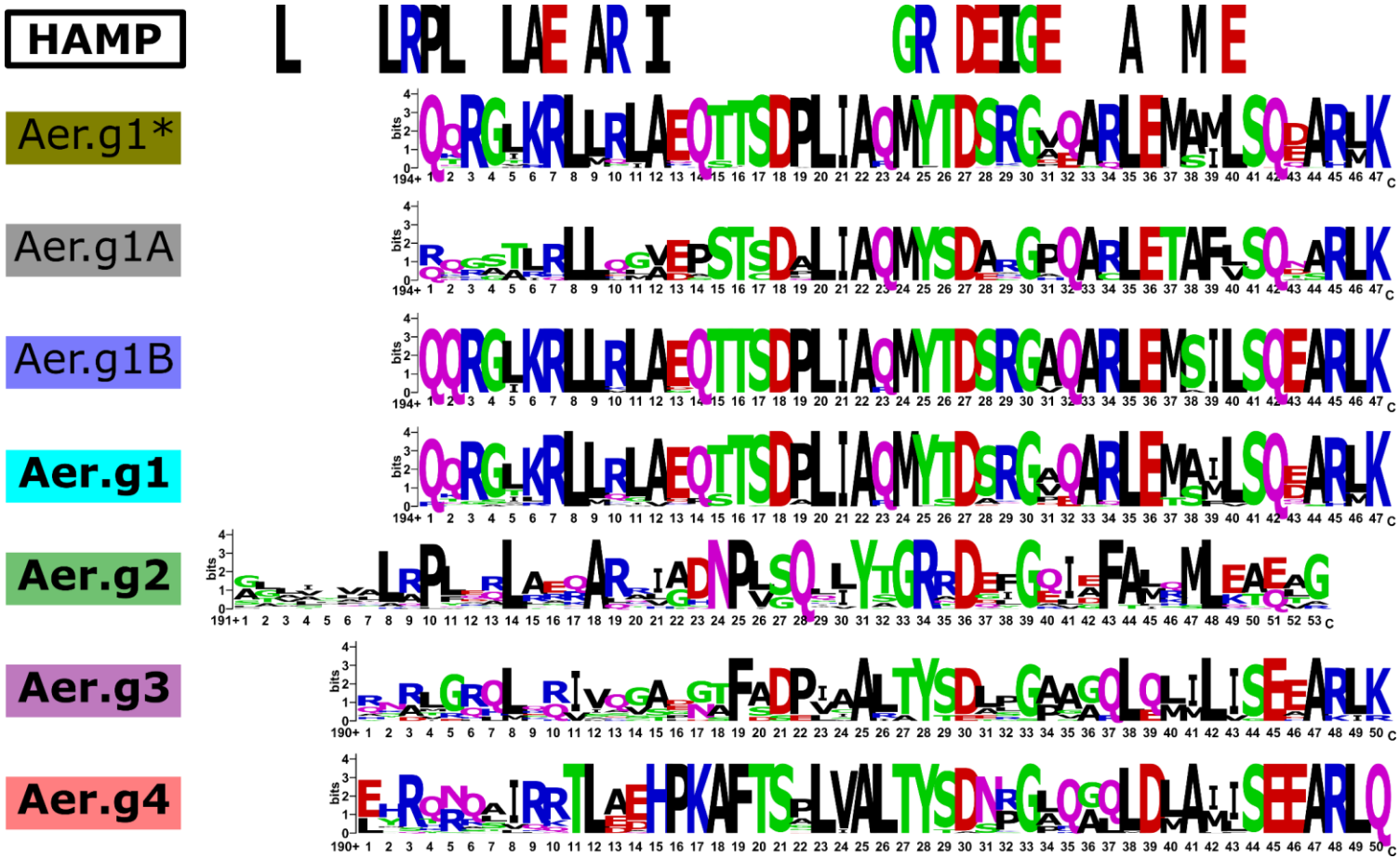
Supplementary Figure 12: Amino acid conservation of the cytoplasmic domain of Aer for all intergroup comparisons, and weblogos for each group. Conservation values were smoothed by taking the average of a position and the 3 preceding and following positions. The cytoplasmic domain is divided into 20 heptads, starting from the central glutamate counting outwards in each direction. Shading above the weblogos indicates the region (heptads 1-4, signaling; heptads 5-12, flexible bundle; heptads 13-20, methylation).



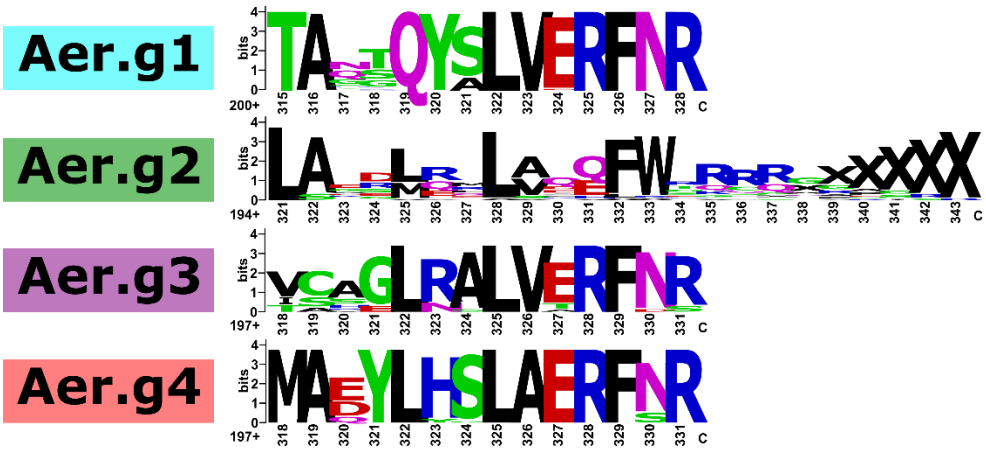
Supplementary Figure 13: Multi-relief scores of the cytoplasmic domain of Aer for all intergroup comparisons, and weblogos for each group. Multi-relief weight values were smoothed by taking the average of a position and the 3 preceding and following positions. Black points are above 0.6, grey below. The cutoff of 0.8 was relaxed due to the smoothing. Black regions indicate group unique regions. The cytoplasmic domain is divided into 20 heptads, starting from the central glutamate counting outwards in each direction. Shading above the weblogos indicates the region (heptads 1-4, signaling; heptads 5-12, flexible bundle; heptads 13-20, methylation).



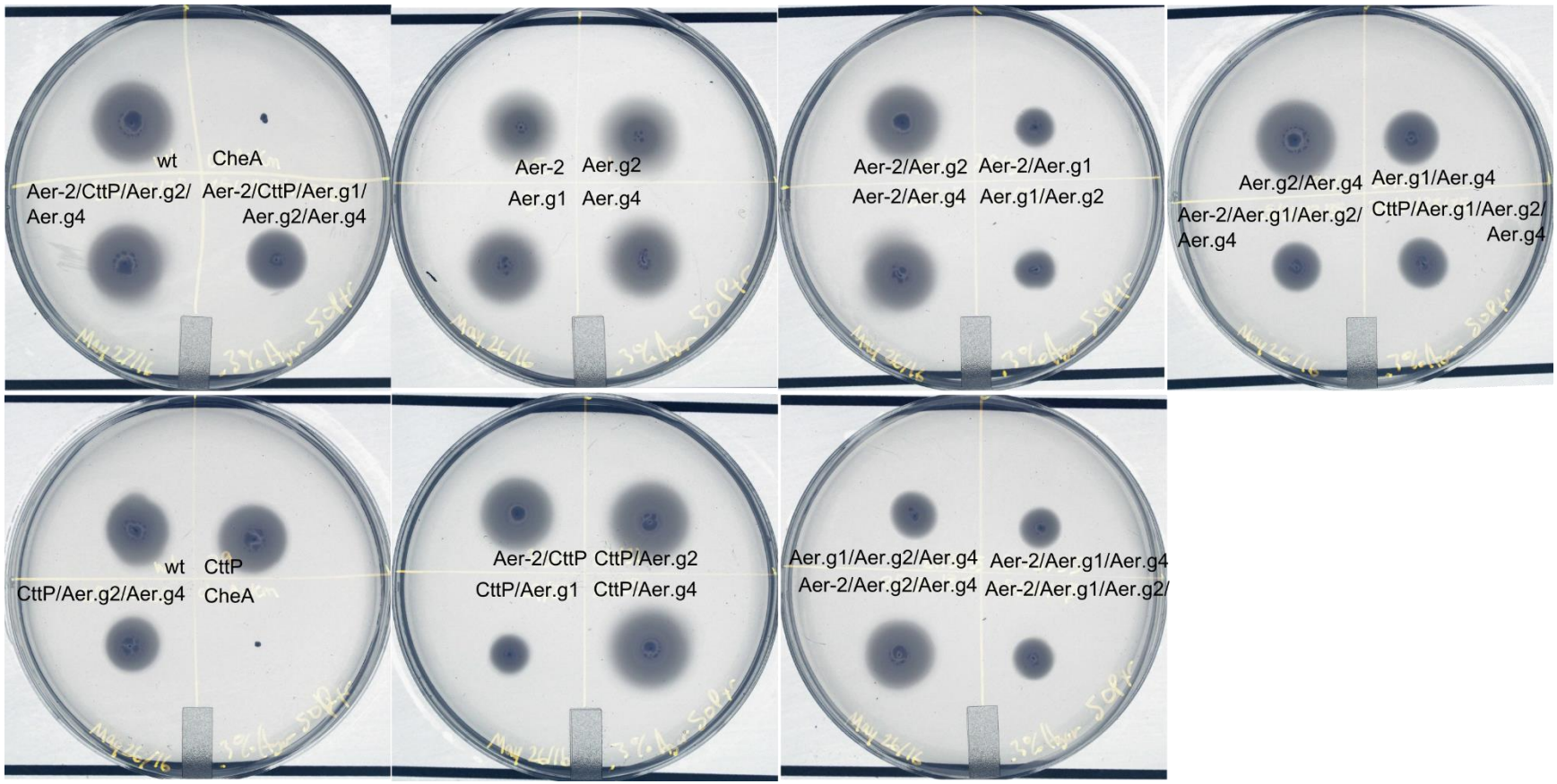
Supplementary Figure 14: Weblogos of each Aer group's PAS domain. Aer.g1* indicates Aer.g1 sequences not including Aer.g1A and Aer.g1B. Aer.g1 includes all Aer.g1 sequences including Aer.g1A and Aer.g1B. Characteristic PAS domain residues were obtained from the SMART database.



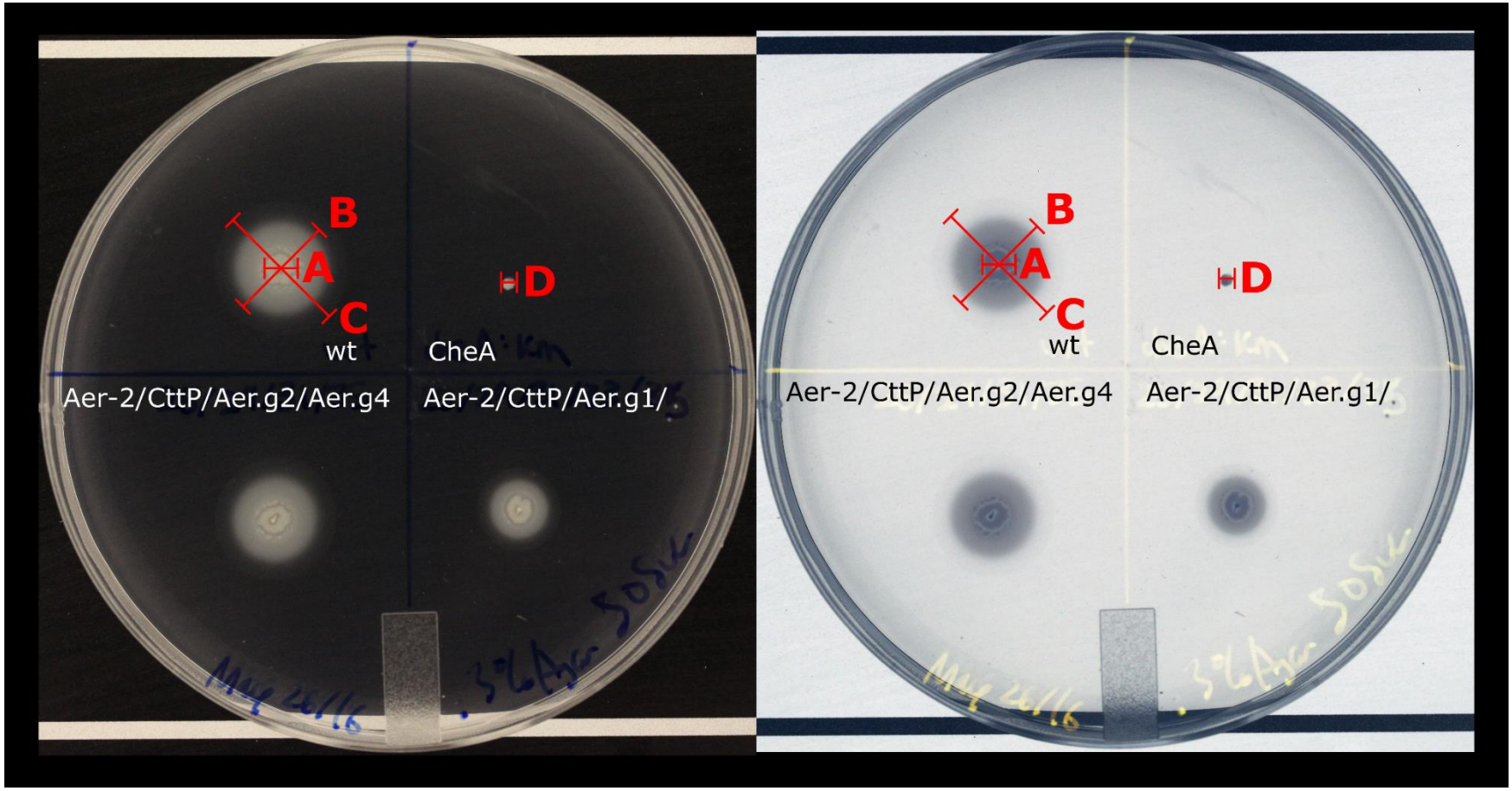
Supplementary Figure 15: Weblogos of Aer group specific region in between end of transmembrane domain and start of cytoplasmic domain heptads. Start of cytoplasmic domain was based on the number of matching heptads on either side of the central glutamate residue (Aer.g1, 388; Aer.g2 394; Aer.g3, 394; Aer.g4 391). The end of the transmembrane domain was determined by submitting the consensus sequence of each group to THMMR. Weblogos were aligned based on the shared aspartate residue (Aer.g1, 221; Aer.g2 227; Aer.g3, 220; Aer.g4 220). Characteristic HAMP domain residues were obtained from the SMART database. This figure is the same as Figure 5, only it includes the Aer.g1 subdivisions. Aer.g1* indicates Aer.g1 sequences not including Aer.g1A and Aer.g1B. Aer.g1 includes all Aer.g1 sequences including Aer.g1A and Aer.g1B. This figure is equivalent to Figure 5, but includes the subdivisions of Aer.g1.



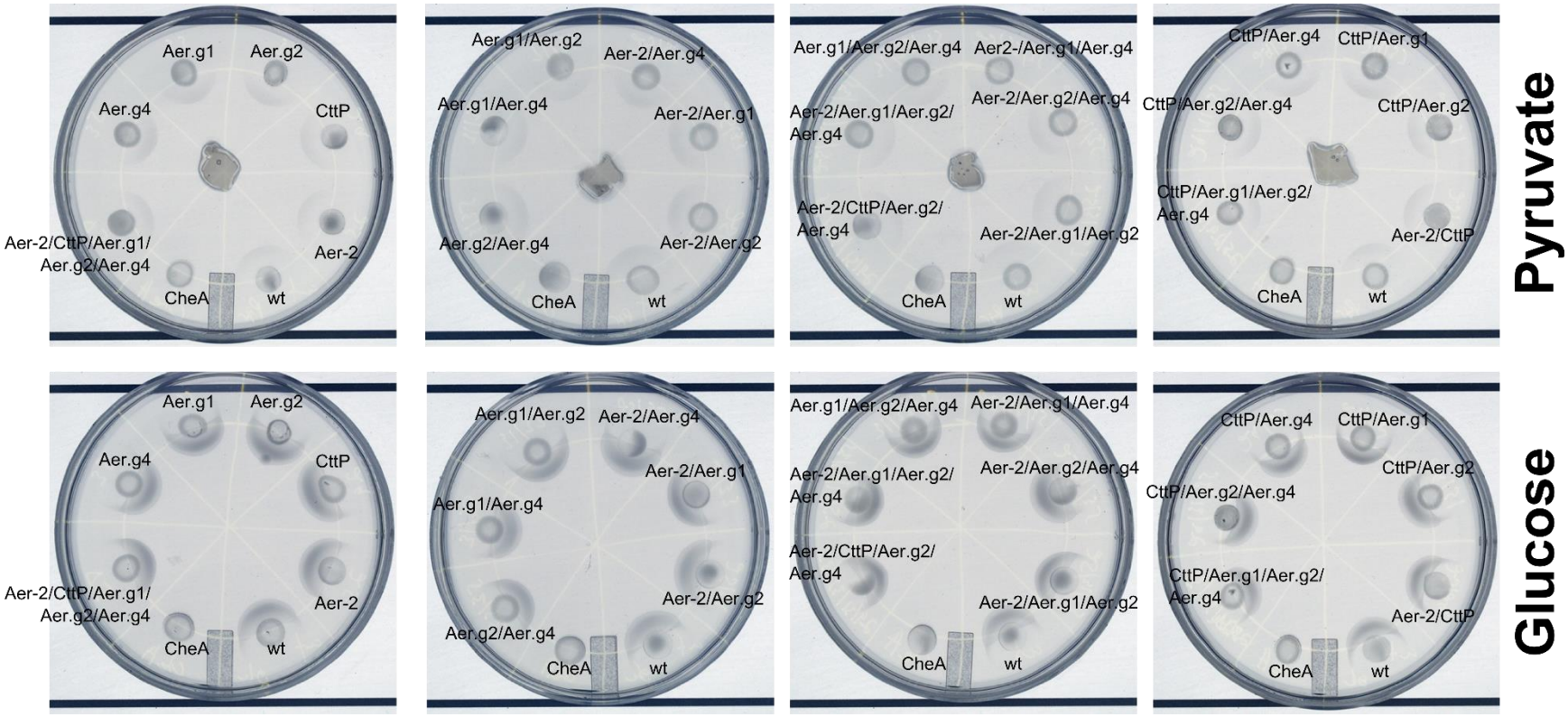
Supplementary Figure 16: Weblogos of Aer C-terminal region. Except for Aer.g2, this constitutes a zoomed in region shown in Supplementary Figure 12. X denotes NO amino acid in that position.



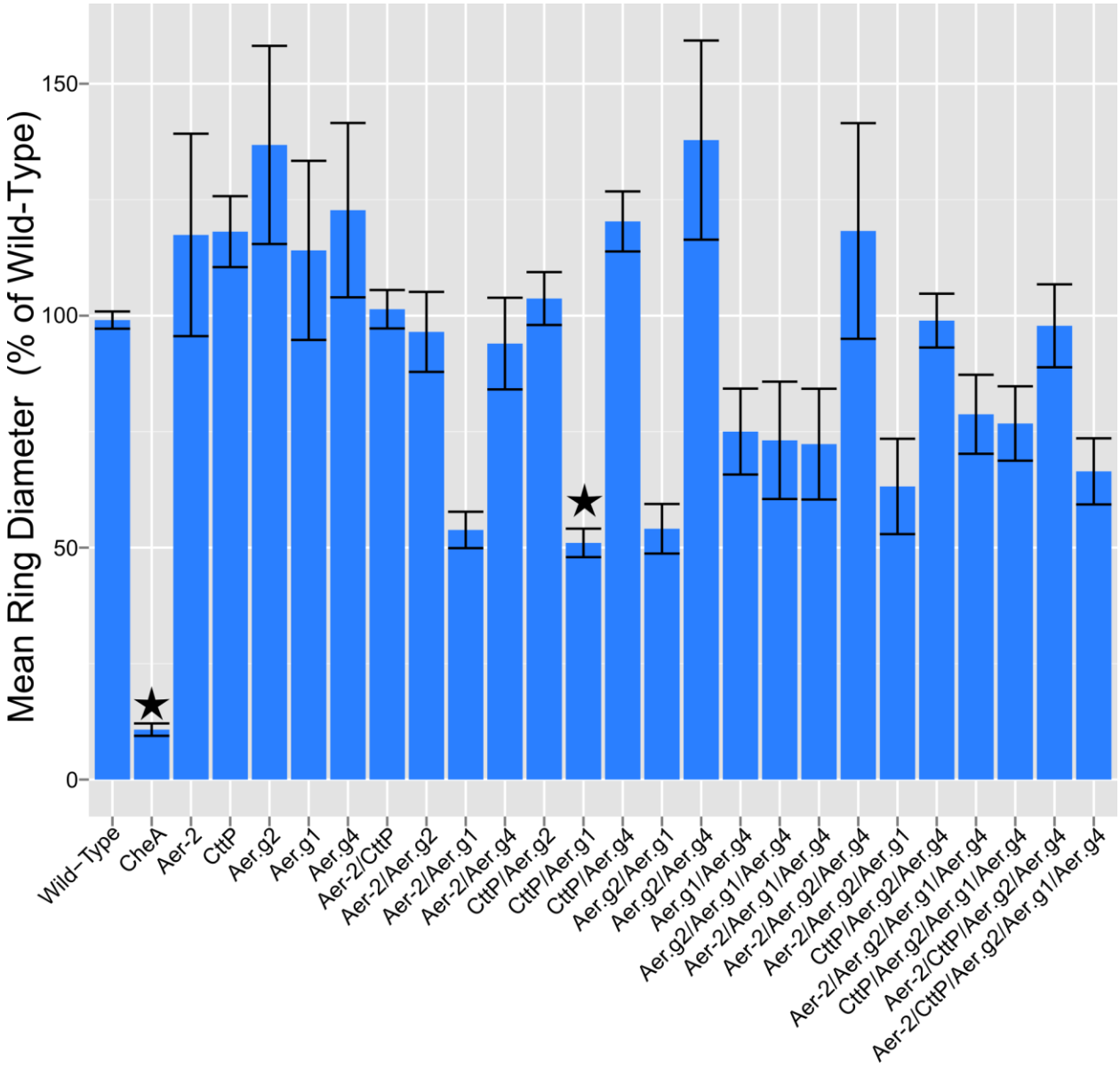
Supplementary Figure 17: Photographs of energy-taxis swim plates of *P. pseudoalcaligenes* strains with deletions in *aer-2*, *cttP* and *aer* homologs after 24h growth at 30°C in 50mM pyruvate. Colours have been inverted to emphasize contrast.



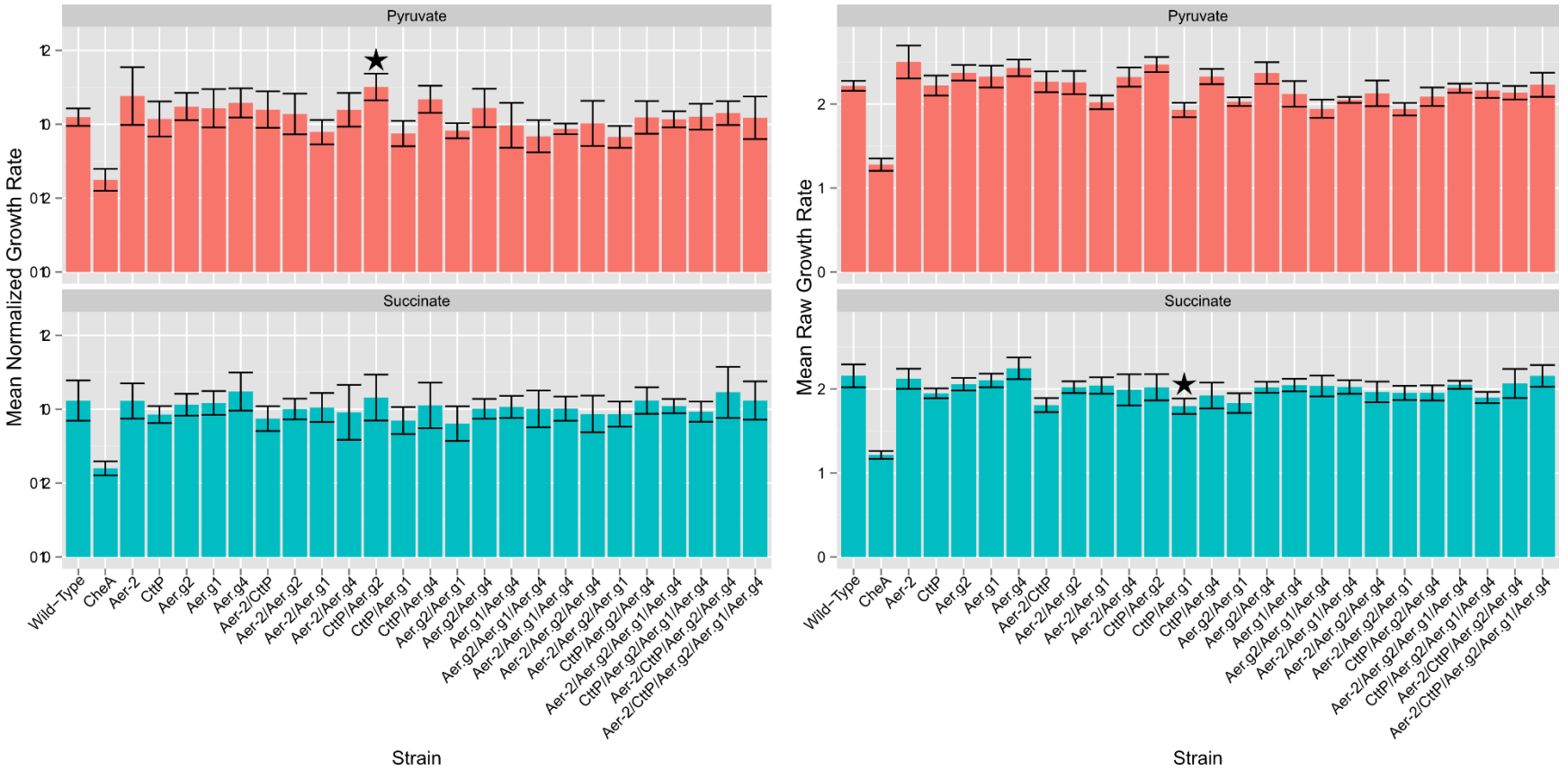
Supplementary Figure 18: Diagram showing how energy-taxis phenotypes were measured. Left is the actual image, right is the same image with colours inverted. This example plate shows the growth of 4 strains after 24h in minimal salts medium, 0.3% agar, 50mM succinate. Strains were inoculated into the plate on a sterile needle. Three different horizons are visible: the edge of dense growth (A), the intermediate colony edge (B), the edge of swimming cells (C). In the CheA mutant, there is no separation of horizons as the strain is non chemotactic (D). The diameter as defined by C that crossed through the inoculation center was used to measure the energy taxis diameter discussed in further figures.



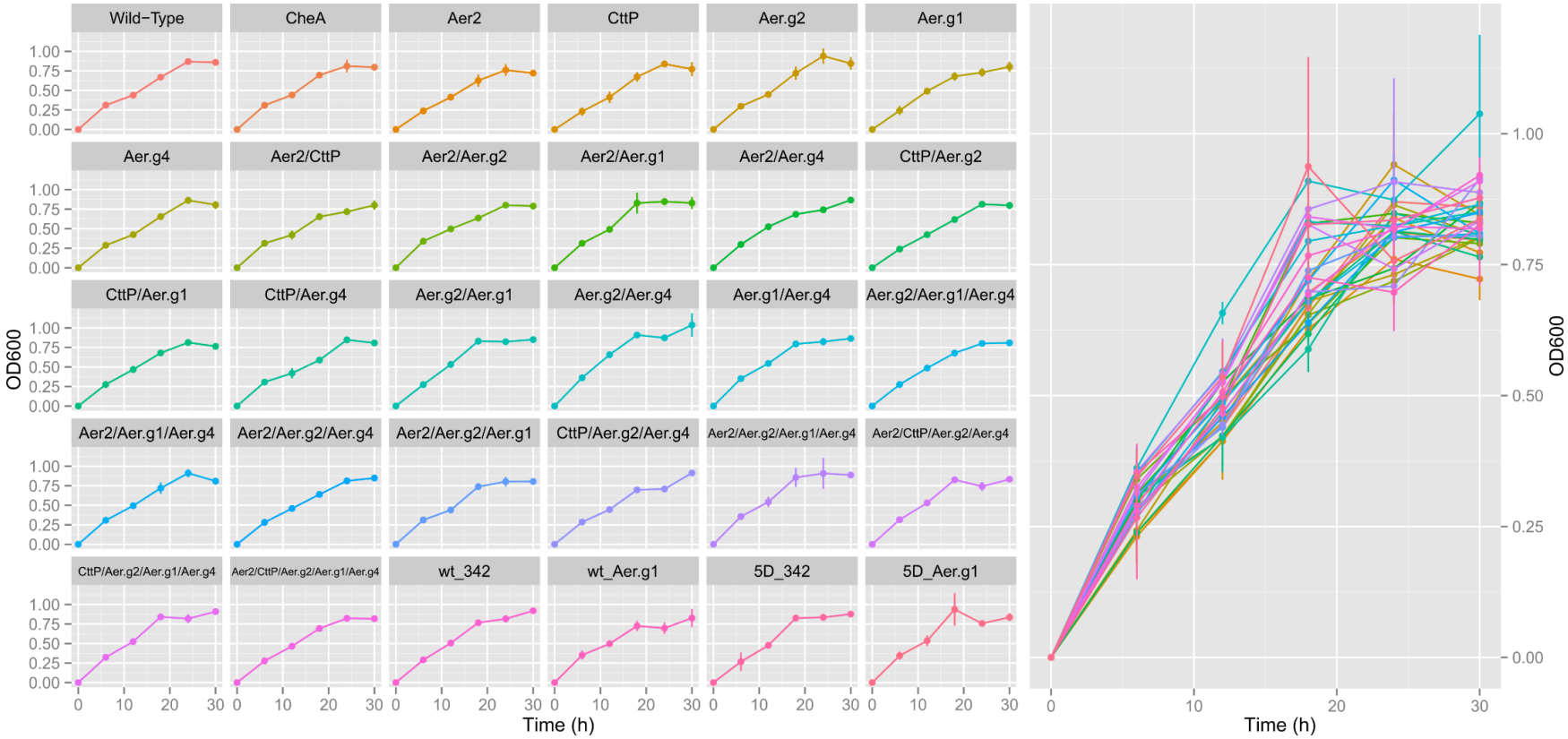
Supplementary Figure 19: Chemotaxis swim assays of *P. pseudoalcaligenes* strains with deletions in *aer-2, cttP* and *aer* homologs. Strains were grown overnight, concentrated then spotted on minimal salts plates containing 0.3% agar. Either 50mM pyruvate in 1.5% agar or crystals of glucose were placed in the centre of the plates. Photographs were taken after 24h. Colours have been inverted to emphasize contrast.



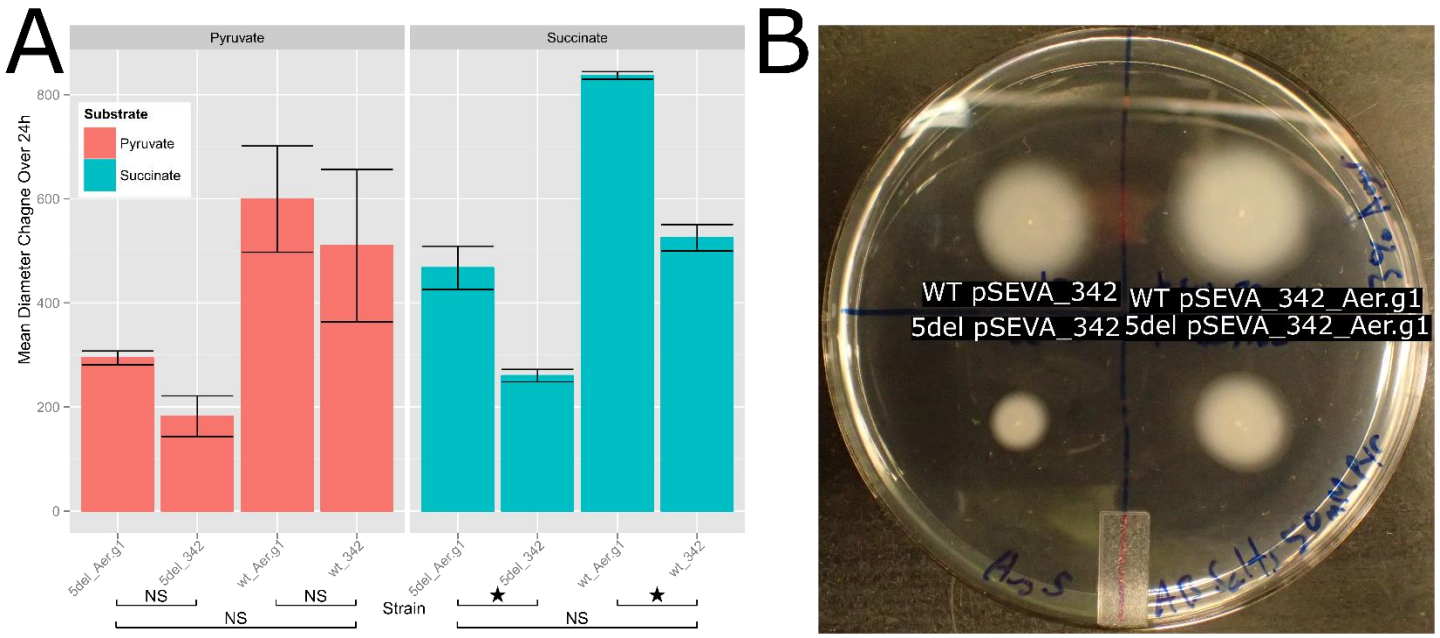
Supplementary Figure 20: Normalized energy-taxis growth diameters in 50mM succinate of strains of *P. pseudoalcaligenes* KF707 with deletions of *aer-2, cttP* and *aer* homologs. Bars indicate the average growth diameter, normalized to the wild-type, at both 24h and 48h from at least 3 experimental replicates. Wild-type strains were normalized to the mean of technical replicates within each experiment. Error bars indicate standard error. Stars indicate significant differences from the wild-type based on Tukey's Honest Significant Differences test with a confidence value of 0.95. The *cheA::KmR* mutant was not grown with antibiotic present.



Supplementary Figure 21: Mean raw and normalized growth rates of energy-taxis swim-diameters of strains of *P. pseudoalcaligenes* KF707 with deletions of *aer-2, cttP* and *aer* homologs. Growth rates were calculated by dividing the diameter at 48h by the diameter at 24h. Normalized rates were calculated from diameters that were normalized to the wild-type for each experiment at each time point. Stars indicate significant differences from the corresponding wild-type according to Tukey's Honest Significant Differences test. Values represent results from at least 3 experimental replicates.



Supplementary Figure 22: Growth *P. pseudoalcaligenes* KF707 strains with deletions of Aer homologs, CttP and Aer-2. Growth was performed in a microtiter plate in minimal salts medium with 10mM pyruvate as the growth substrate. Growth of the wild-type and quintuple mutant (Aer-2/CttP/Aer.g1/Aer.g2/Aer.g4) with the empty vector pSEVA342 and pSEVA342_Aer.g1 was also assayed. Values represent the average of 2 biological replicates from 2 experimental replicates (4 total replicates). Lines at points indicates standard error.



Supplementary Figure 23: Mean change of energy-taxis diameters over 24h of strains of *P. pseudoalcaligenes* KF707 wild-type and quintuple mutant strains carrying Aer.g1 complementation plasmids. Change is the difference between 24h and 48h (A). Stars indicate significant differences according to Tukey's Honest Significant Differences test. Values represent results from 2 experimental replicates. Strains were carrying either pSEVA_342 (empty vector control) or p_SEVA_342 with Aer.g1. 5del indicates quintuple mutant with deletion of Aer2/CttP/Aer.g2/Aer.g1/Aer.g4. Photograph in bottom right was taken after 24h of growth (B).

Supplementary Table 1: Genes immediately adjacent to *aer* genes and presence of mobile elements. See supplementary data.

Supplementary Table 2: Frequency of association of mobile elements within 5kb of *aer* genes. Summarized from supplementary Table 1.

Aer Group	Repeats	Transposase	Integrase
aer.g1	20/73	2/73	0/73
aer.g2	4/26	1/26	0/26
aer.g3	1/16	1/16	1/16
aer.g4	4/27	2/27	1/27
aer.g5	0/2	0/2	0/2
Total	29/144	6/144	2/144

Supplementary Table 3: Tukey Honest Significant Differences results comparing differences in normalized energy-taxis diameter in pyruvate or succinate plates. HSD test compared all pairs of strains, only comparisons to the wild-type are presented here. P values were computed using a 0.95 confidence value, those below 0.05 were taken as significant.

Comparison	Pyruvate Difference	Pyruvate Adjusted p Value	Succinate Difference	Succinate Adjusted p Value
Wild-Type-Aer.g1	0.044557106	1	-0.15022	1
Wild-Type-Aer.g1/Aer.g4	0.316929667	**0.001584542**	0.240355	0.995568
Wild-Type-Aer.g2	-0.130260185	0.994474052	-0.37755	0.470077
Wild-Type-Aer.g2/Aer.g1	0.416623963	**6.99E-07**	0.449853	0.123956
Wild-Type-Aer.g2/Aer.g1/Aer.g4	0.41531338	**7.83E-07**	0.259256	0.985363
Wild-Type-Aer.g2/Aer.g4	-0.107657177	0.999855723	-0.38801	0.404151
Wild-Type-Aer.g4	-0.080272111	0.999999891	-0.23695	0.996521
Wild-Type-Aer2	0.06079056	1	-0.18355	0.99998
Wild-Type-Aer2/Aer.g1	0.423034811	**3.99E-07**	0.452405	0.116907
Wild-Type-Aer2/Aer.g1/Aer.g4	0.348397359	0.000171831	0.267403	0.977193
Wild-Type-Aer2/Aer.g2	-0.028743339	1	0.025494	1
Wild-Type-Aer2/Aer.g2/Aer.g1	0.424269068	**3.58E-07**	0.35864	0.594415
Wild-Type-Aer2/Aer.g2/Aer.g1/Aer.g4	0.358356796	**8.13E-05**	0.203102	0.999816
Wild-Type-Aer2/Aer.g2/Aer.g4	0.172011243	0.810645941	-0.19209	0.999944
Wild-Type-Aer2/Aer.g4	0.00831534	1	0.050637	1
Wild-Type-Aer2/CttP	-0.05951297	0.999999999	-0.02338	1
Wild-Type-Aer2/CttP/Aer.g2/Aer.g1/Aer.g4	0.396996279	3.72E-06	0.326159	0.792546
Wild-Type-Aer2/CttP/Aer.g2/Aer.g4	0.063072257	0.999999997	0.012354	1
Wild-Type-CheA	0.87948192	0	0.88287	**3.16E-10**
Wild-Type-CttP	-0.139877304	0.983128982	-0.19063	0.999953
Wild-Type-CttP/Aer.g1	0.459331863	**4.31E-11**	0.480318	**0.011177**
Wild-Type-CttP/Aer.g2	-0.025195491	1	-0.04643	1
Wild-Type-CttP/Aer.g2/Aer.g1/Aer.g4	0.38554092	**1.60E-07**	0.22296	0.991638
Wild-Type-CttP/Aer.g2/Aer.g4	0.015579387	1	0.001138	1
Wild-Type-CttP/Aer.g4	-0.145041436	0.935787346	-0.21263	0.998431

Supplementary Table 4: p values from Tukey HSD results comparing differences in raw and normalized energy-taxis diameter growth rates in pyruvate or succinate plates. Growth rates were obtained by dividing the raw or normalized diameter at 48h by the value at 24h. HSD test compared all pairs of strains, only comparisons to the wild-type are presented here. P values were computed using a 0.95 confidence value, those below 0.05 were taken as significant.

Comparison	Pyruvate Normalized	Succinate Normalized	Pyruvate Raw	Succinate Raw
Wild-Type-Aer.g1	1	1	1	1
Wild-Type-Aer.g1/Aer.g4	1	1	1	1
Wild-Type-Aer.g2	1	1	0.999	1
Wild-Type-Aer.g2/Aer.g1	0.997992	0.194316	0.96148	0.37878
Wild-Type-Aer.g2/Aer.g1/Aer.g4	0.826779	1	0.332046	1
Wild-Type-Aer.g2/Aer.g4	1	1	0.999447	1
Wild-Type-Aer.g4	0.999305	0.999923	0.894703	0.999999
Wild-Type-Aer2	0.791526	1	0.305462	1
Wild-Type-Aer2/Aer.g1	0.990515	1	0.937616	1
Wild-Type-Aer2/Aer.g1/Aer.g4	0.999856	1	0.99263	1
Wild-Type-Aer2/Aer.g2	1	1	1	1
Wild-Type-Aer2/Aer.g2/Aer.g1	0.776794	0.994759	0.290882	0.997895
Wild-Type-Aer2/Aer.g2/Aer.g1/Aer.g4	1	1	1	1
Wild-Type-Aer2/Aer.g2/Aer.g4	1	0.995619	1	0.999439
Wild-Type-Aer2/Aer.g4	1	0.99969	1	0.999984
Wild-Type-Aer2/CttP	1	0.516674	1	0.066568
Wild-Type-Aer2/CttP/Aer.g2/Aer.g1/Aer.g4	1	1	1	1
Wild-Type-Aer2/CttP/Aer.g2/Aer.g4	1	0.999829	1	1
Wild-Type-CheA	**0**	**0**	**0**	**0**
Wild-Type-CttP	1	0.990882	1	0.996631
Wild-Type-CttP/Aer.g1	0.866421	0.259124	0.070324	**0.039548**
Wild-Type-CttP/Aer.g2	**0.03075**	1	0.416662	1
Wild-Type-CttP/Aer.g2/Aer.g1/Aer.g4	1	0.999082	1	0.712737
Wild-Type-CttP/Aer.g2/Aer.g4	1	1	0.999329	0.986805
Wild-Type-CttP/Aer.g4	0.926885	1	0.999998	0.939527

Supplementary Table 5: p values from Tukey HSD results comparing differences in energy-taxis diameters, diameter changes and diameter growth rates in pyruvate or succinate plates for complementation strains. 5del indicates deletion of Aer2/CttP/Aer.g2/Aer.g1/Aer.g4. Growth rates were obtained by dividing the raw or normalized diameter at 48h by the value at 24h. Amount of growth was obtained by subtracting the 24h diameter from the value at 48h. HSD test compared all pairs of strains. P values were computed using a 0.95 confidence value, those below 0.05 were taken as significant (bold).

Parameter	Time	Comparison	Succinate Difference	Succinate Adjusted p Value	Pyruvate Difference	Pyruvate Adjusted p Value
Diameter	24	5del_342-5del_33	-57.607	0.920	-59.515	0.644
		wt_Aer.g1-5del_Aer.g1	254.831	0.157	119.757	0.204
		wt_342-5del_Aer.g1	149.687	0.463	93.301	0.346
		wt_Aer.g1-5del_342	312.438	0.089	179.272	0.067
		wt_342-5del_342	207.294	0.256	152.816	0.108
		wt_342-wt_Aer.g1	-105.144	0.691	-26.456	0.943
Diameter	48	5del_342-5del_Aer.g1	-264.812	0.196	-171.301	0.755
		wt_Aer.g1-5del_Aer.g1	624.510	**0.014**	425.043	0.199
		wt_342-5del_Aer.g1	207.487	0.331	309.349	0.382
		wt_Aer.g1-5del_342	889.321	**0.004**	596.343	0.079
		wt_342-5del_342	472.299	**0.036**	480.649	0.146
		wt_342-wt_Aer.g1	-417.023	0.054	-115.694	0.900
Growth	NA	5del_342-5del_Aer.g1	-207.205	**0.015**	-111.786	0.824
		wt_Aer.g1-5del_Aer.g1	369.679	**0.002**	305.286	0.228
		wt_342-5del_Aer.g1	57.800	0.461	216.048	0.441
		wt_Aer.g1-5del_342	576.884	**0.000**	417.072	0.102
		wt_342-5del_342	265.005	**0.006**	327.834	0.193
		wt_342-wt_Aer.g1	-311.879	**0.003**	-89.238	0.897
Rate	NA	5del_342-5del_Aer.g1	-0.422	0.265	-0.235	0.745
		wt_Aer.g1-5del_Aer.g1	0.134	0.892	0.430	0.362
		wt_342-5del_Aer.g1	-0.224	0.674	0.284	0.639
		wt_Aer.g1-5del_342	0.556	0.136	0.665	0.136
		wt_342-5del_342	0.198	0.743	0.519	0.249
		wt_342-wt_Aer.g1	-0.358	0.365	-0.146	0.915

Supplementary Table 6: Plasmids used for energy-taxis experiments in *Pseudomonas pseudoalcaligenes* KF707.

Plasmid	Description	Reference
pRK2013	KmR ori colE1 RK2-Mob$^+$ RK2-Tra$^+$	(2)
pG19II	Gm^R, *sacB*, *lacZ*, cloning vector, conjugative plasmid	(3)
pG19II-*Δaer-2*	Gm^R, *sacB*, *lacZ*, *aer-2* deletion construct	This Study
pG19II-*ΔcttP*	Gm^R, *sacB*, *lacZ*, *cttP* deletion construct	This Study
pG19II-*Δaer.g1*	Gm^R, *sacB*, *lacZ*, *aer.g1* deletion construct	This Study
pG19II-*Δaer.g2*	Gm^R, *sacB*, *lacZ*, *aer.g2* deletion construct	This Study
pG19II-*Δaer.g4*	Gm^R, *sacB*, *lacZ*, *aer.g4* deletion construct	This Study
pSEVA324	Cm^R, pR01600/ColE1, *lacZα*-pUC19	(4)
pSEVA324-*aer.g1*	Cm^R, pSEVA342 with *aer.g1* in the MCS	This Study

Supplementary References

1. **Schultz J**, **Copley RR**, **Doerks T**, **Ponting CP**, **Bork P**. 2000. SMART: a web-based tool for the study of genetically mobile domains. Nucleic Acids Res **28**:231–4.

2. **Figurski DH**, **Helinski DR**. 1979. Replication of an origin-containing derivative of plasmid RK2 dependent on a plasmid function provided in trans. Proc Natl Acad Sci U S A **76**:1648–52.

3. **Maseda H**, **Sawada I**, **Saito K**, **Uchiyama H**, **Nakae T**, **Nomura N**. 2004. Enhancement of the mexAB-oprM efflux pump expression by a quorum-sensing autoinducer and its cancellation by a regulator, MexT, of the mexEF-oprN efflux pump operon in Pseudomonas aeruginosa. Antimicrob Agents Chemother **48**:1320–8.

4. **Silva-Rocha R**, **Martínez-García E**, **Calles B**, **Chavarría M**, **Arce-Rodríguez A**, **De Las Heras A**, **Páez-Espino AD**, **Durante-Rodríguez G**, **Kim J**, **Nikel PI**, **Platero R**, **De Lorenzo V**. 2013. The Standard European Vector Architecture (SEVA): A coherent platform for the analysis and deployment of complex prokaryotic phenotypes. Nucleic Acids Res **41**:D666–D675.