Aer is a family of Energy-Taxis Receptors in *Pseudomonas* which also involves Aer-2 and CttP

Sean C. Booth and Raymond J. Turner

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

#Address correspondence to Sean C. Booth, scbooth@ucalgary.ca

Abstract

Energy-taxis by *Pseudomonas* species allows swimming towards optimal environments for generating cellular energy, precluding the need to sense carbon sources and oxygen. Aerotaxis, which has been taken to mean swimming towards oxygen or a source of metabolizable carbon has been attributed to the methyl-accepting chemotaxis protein Aer. In *P. aeruginosa* McpB (Aer-2) also contributes and may directly sense oxygen whereas in *P. putida* the second of three Aer energy-taxis sensor homologs, 'Aer2' is the key receptor. Here we sought to disentangle the ambiguity between aerotaxis and energy-taxis and between McpB (Aer-2) and Aer2. Phylogenetic characterization of sequences from a wide range of *Pseudomonas* species revealed that Aer homologs have been duplicated and horizontally transferred within the genus but can be divided into 5 groups. Phenotypic characterization in *P. pseudoalcaligenes* KF707, which has 3 Aer homologs and McpB showed that all homologs as well as McpA (CttP) and McpB (Aer-2) contribute to energy-taxis, though the most ancestral Aer homolog is key but not essential to this behaviour. As McpB (Aer-2) is rare in the genus, these results indicate that energy-taxis and aerotaxis are linked but distinct behaviours and that energy-taxis in *Pseudomonas* involves more genes than previously thought.

Importance

Energy-taxis is a behaviour in *Pseudomonas* that allows it to swim towards metabolizable carbon sources by sensing the cellular energy they generate rather than the actual compound being metabolized. This ability has been attributed to the receptor 'Aer', of which some species of *Pseudomonas* have multiple homologs. There is confusion in the field as functional homologs have been named Aer2 which clashes with a previously named, completely distinct receptor called Aer-2, which is thought to be an oxygen receptor. Swimming towards oxygen is called aerotaxis and

30 has been conflated with energy-taxis. Since both are useful behaviours this study aimed to sort out
31 the differences between them, and the differences between Aer2 and Aer-2.

32 **Introduction**

33 Chemotaxis, the ability to sense and swim along chemical gradients, is a widespread and
34 important behaviour in bacteria (1). Canonically it functions by extracellular compounds binding
35 to membrane-bound methyl-accepting chemotaxis proteins (MCPs) causing a phosphorylation
36 signal cascade through CheA and CheY to alter the direction of flagellar rotation (1), allowing the
37 cell to direct swimming through concentration gradients. The first MCPs were characterized in *E.*
38 coli but now many more have been described, particularly in *Pseudomonas* (2). While receptors
39 for specific ligands are being identified, many through a clever high-throughput approaches (3),
40 our understanding of some 'characterized' MCPs remains incomplete. The receptors in
41 *Pseudomonas* involved in aerotaxis where energy-taxis have been identified, but questions remain
42 about their exact function and whether the two processes are actually distinct.

43 Chemotaxis is generally used as a catchall to describe any kind of directed bacterial motion
44 towards or away from a chemical, but for the purposes of this study chemotaxis, aerotaxis and
45 energy-taxis must be more strictly defined. Chemotaxis refers to directed swimming mediated by
46 binding of a ligand either directly or through a periplasmic binding protein to an MCP.
47 Distinguishing between aerotaxis and energy-taxis is more important as these have been
48 traditionally conflated, energy-taxis being a relatively new term. Here aerotaxis will be defined as
49 directed swimming towards oxygen through direct sensing of O_2, and energy-taxis as directed
50 swimming towards a more favourable environment through indirect sensing of energy production
51 within the cell. Aerotaxis was among the first bacterial behaviours ever described as both
52 Engelmann and Beijerinck described the aggregation of bacterial cells near sources of oxygen (4,

5). When the MCP responsible for this phenomenon was discovered in *E. coli*, it was given the name Aer (6). While it does indeed mediate taxis towards oxygen, the mechanism was later found not to involve direct binding of oxygen (7). Instead, Aer has an FAD cofactor which when reduced allows the bacterium to sense energy generation within the cell, meaning Aer is actually an energy-taxis receptor. The distinction between energy and aerotaxis became important when two genes in *P. aeruginosa* were found to be involved in 'aerotaxis', one a homolog of Aer, the second a distinct cytoplasmic MCP which was given the name Aer-2 (8), though it is also still called McpB. Further issues were raised when 3 homologs of Aer were found in *P. putida* (9), and the most abundant and only functional homolog was given the name 'Aer2'. Biochemical characterization of Aer-2 (also called McpB or TlpG) showed that it binds gases including O_2 but functional reports disputed its role in 'aerotaxis' (10). Thus despite multiple investigations it remains unclear which genes are responsible for aerotaxis and energy-taxis in *Pseudomonas*, and whether they are distinct processes. Here we sought answers to these questions using a combination of comparative genomics and phenotypic characterization of mutants.

In *Pseudomonas,* Aer was first characterized in *P. putida* PRS2000 (11), then in *P. aeruginosa* PA01 (8), then in *P. putida* KT2440 (9) and F1 (12) as an 'aerotaxis' receptor, using a variety of methods. Unfortunately, it is difficult to discern a difference between energy-taxis from aerotaxis from the experimental approach used in such studies. Despite the observation that *P. putida* had 3 homologs whose functionality had to be determined by process of elimination, compared to *P. aeruginosa's* 1, no comparison of the genes or gene products was made. This raises the question of how homologs are distributed in the genus, which ones are functional in other species, and most importantly why do they possess multiple similar copies of the same chemoreceptor? Aer-2 was originally implicated with Aer as an aerotaxis receptor and while the

76 experimental evidence characterizing CttP as a TCE receptor is clear (13), yet it seems unlikely
77 that this is the natural function of the protein. Due to its appearance as the first gene in the *che2*
78 operon, which also contains *mcpB* (*aer-2*) we hypothesized that both may have a role in energy-
79 taxis. To investigate these questions, Aer sequences were obtained from 65 *Pseudomonas* species
80 as well as the presence/absence of Aer-2 and CttP. A phylogeny of Aer was built providing insight
81 into the distribution in the genus and allowing the 'Aer' family to be defined. Only one species, *P.*
82 *pseudoalcaligenes* KF707 had the surprising feature of possessing 3 Aer homologs, Aer-2 and
83 CttP making it the ideal candidate to investigate if these receptors had related functions. Deletion
84 mutants were generated revealing that in KF707, all 5 of these genes contributed to energy-taxis,
85 though the ancestral Aer homolog was pivotal. Together these results provide a definition of the
86 Aer energy-taxis receptor as a family with varied distribution in *Pseudomonas* and implicate Aer-
87 2 and CttP in this behaviour.
88

89 **Results**

90 65 *Pseudomonas* species were selected for this analysis. Species with completely
91 sequenced genomes were selected, but for those with many strains only a few representatives that
92 have been highly studied were included. Species with incomplete (draft) sequences were also
93 included in an attempt to ensure representation from all major *Pseudomonas* clades, based on
94 works by (32, 33). 144 protein sequences were obtained from the NCBI database using the *P.*
95 *aeruginosa* PA01 Aer sequence (NP_250252.1) as a BLAST (15) query sequence. All hits with
96 >95% sequence coverage were included. These sequences were aligned using COBALT (17) as
97 this alignment algorithm ensures that conserved domains are aligned despite a lack of similarity
98 elsewhere in the sequence. This was ideal for the Aer sequences as, like other MCPs, they are

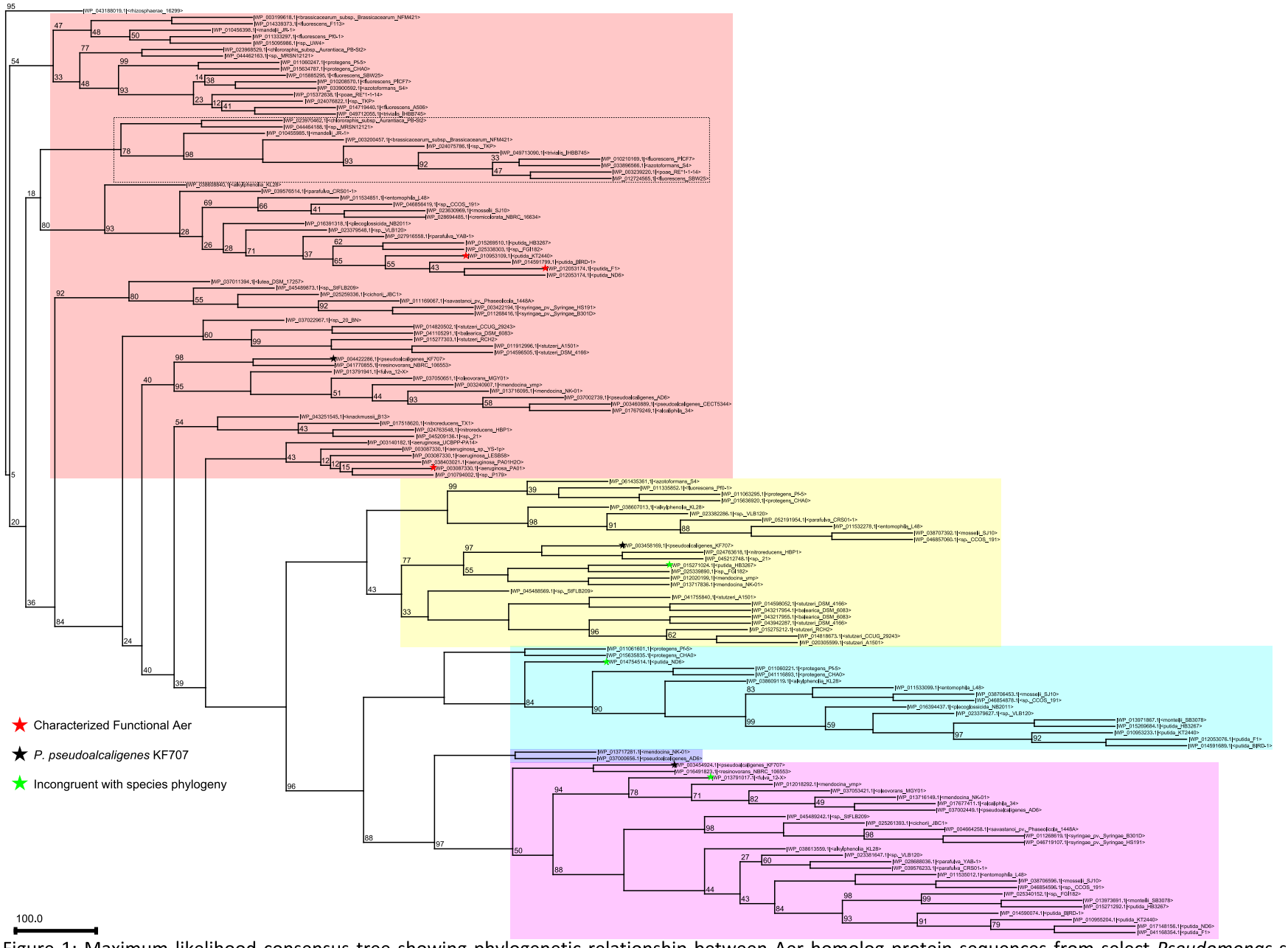

Figure 1: Maximum likelihood consensus tree showing phylogenetic relationship between Aer homolog protein sequences from select *Pseudomonas* species. Colours indicate homolog group: Aer.g1 (red), Aer.g2 (yellow), Aer.g3 (cyan), Aer.g4 (purple) and Aer.g5 (blue). Groups were based on inspection of the alignment used to generate the phylogeny and other analyses, see the text for details. Tree was generated unrooted, then the root placed based on a similar tree that was rooted to the closes non-*Pseudomonas* sequence available. Branch lengths indicate number of AA substitutions per site. Numbers at nodes indicate bootstrap support values from 100 replicates. Dotted box indicates original separation of Aer.g1 into two groups.

conserved in the C-terminal signaling region but less conserved in the N-terminal transmembrane and ligand binding region. Inspection of the alignment revealed numerous unnecessary gaps, which could not be remedied by changing the alignment settings. Instead, these gaps were removed and the regions re-aligned manually. This altered alignment was then re-aligned using MUSCLE (19) to remedy any small incongruities introduced by the manual process. This produced a suitable alignment for determining the phylogeny of the sequences (Figure 1). The Mobyle web platform (21) was used to construct a maximum likelihood phylogeny using PHYML (22). Two phylogenies were generated, one using only the *Pseudomonas* sequences and one which included the closest non-*Pseudomonas* sequence in the NCBI database to be used as a root (WP_027981445.1, unknown gamma proteobacterium L18, (34)). This rooted phylogeny (Supplementary Figure 1) was examined and the location of the root was used to place the root on the unrooted tree. This revealed that the Aer homolog from *P. rhizospherae* 16299 (WP_043188019.1) was more similar to the non-*Pseudomonas* root than to any *Pseudomonas* sequence.

Grouping of Aer Homologs

Inspection of the unrooted, phylogenetically organized alignment implied that there are several sub-families of Aer homologs (Figure 1). Initially, 7 groups were identified based on the alignment and tree branching. These groups were aligned by themselves then analyzed to determine if the group assignments were accurate. Manual inspection and Sequence Harmony / Multi-Relief (SHMR) demonstrated that there were actually only 5 groups. 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 (23). 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 2). AAs with a multi-relief weight score >0.8 were accepted as fitting these criteria (cutoff is based on the recommendations of the SHMR authors (23)). Comparisons of groups where only 10% of AAs fit into this category indicated that these two groups were too similar as most comparisons indicated that 25-35% of AAs should be above the distinction cutoff. Through this process, groups 1 and 2 were combined and groups 3A and 3B were combined. Re-analysis using SHMR showed that all inter-group comparisons had between 25-35% of AAs distinguishing the two groups (Supplementary Figure 3). Throughout this process, group 5 was consistently excluded as it only contains 2 sequences, which are highly divergent from all others. When the genomic context of each *aer* homolog gene was later examined (Figure 4), showing that the genes from the same group shared synteny. This supports the conclusion that there are 5 groups within the Aer family, and they are defined by shared amino acid sequence features and genes they are associated with.

Comparison of Groups

Results from the SHMR analysis for each group comparison were plotted against the entire length of the Aer protein and compared to the overall conservation at each AA position (Figure 2). Using a smoothing function, the regions that were varying between groups were identified and compared to the domain architecture of the Aer protein. These comparisons showed that the regions determining the specificity of each group were the same for all groups, (except for two regions at each extremity of the protein that were specific to Aer.g2) and that these regions were poorly conserved in the overall alignment. As expected, the CheW interface region of the methyl-accepting domain and the Per/Arnt/Sim (PAS) ligand binding domain were well conserved and did not differ between groups, though the CheW interface was far more conserved than the PAS domain. Two additional conserved regions (deemed C1 and C3) were found before and after the

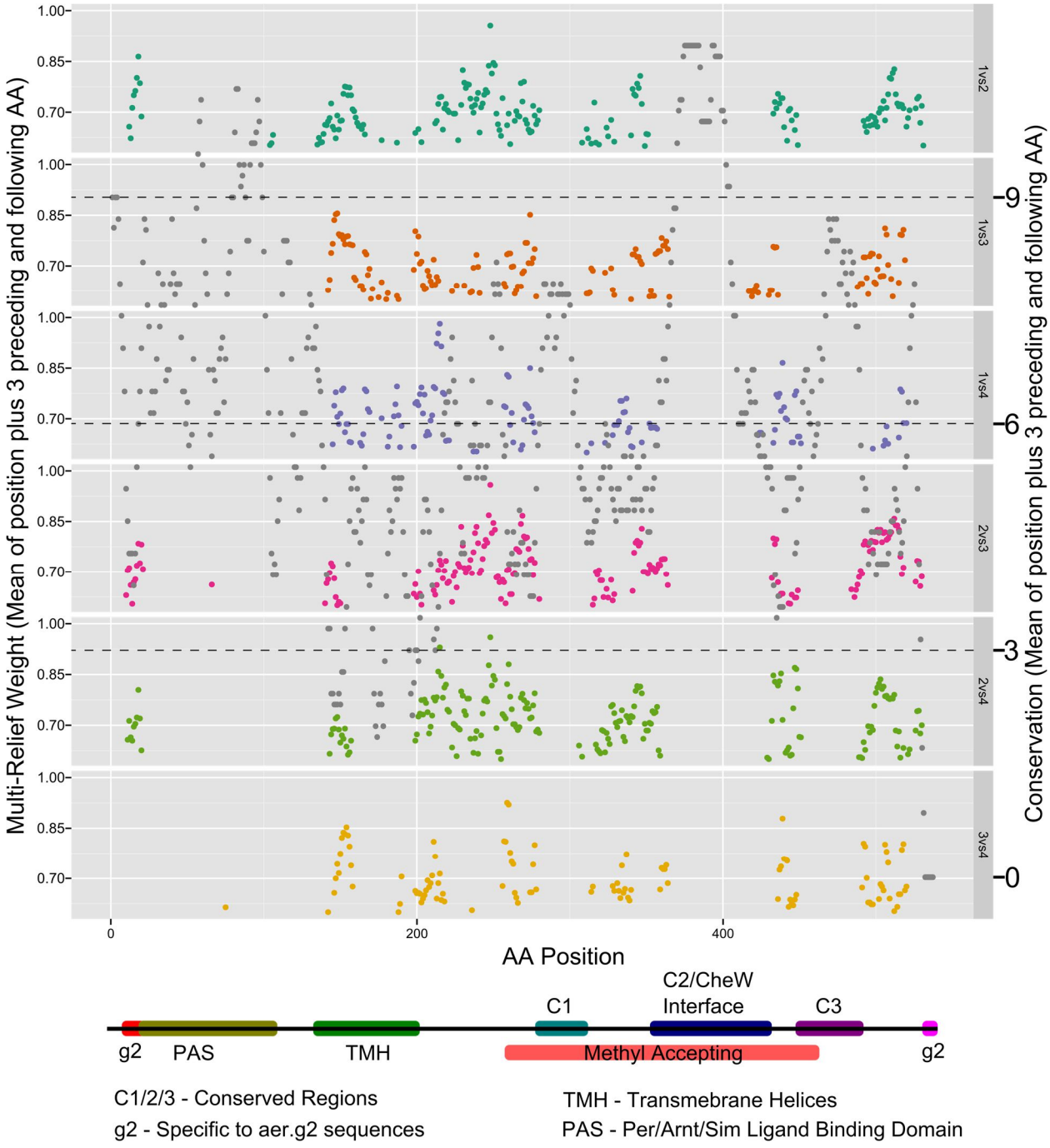

Figure 2: Amino acid conservation and multi-relief weight scores for all positions of Aer and domain architecture of Aer. Conservation was calculated across the entire alignment of all 144 Aer homolog sequences. Multi-relief scores were calculated for each pair-wise comparison of homolog groups. Both values have been smoothed by taking the sum of the 3 proceeding and following AAs for each position. Domain architecture was obtained from the conserved domain database. Regions have been colour coded based on known features, and from features derived from the current study.

158 CheW interface region, each separated by an unconserved, group specific region. The predicted
159 transmembrane domains (AAs ~130-200) were the least conserved region and were not conserved
160 within groups except for the initial portion of these domains (AA ~150). Between this region and
161 the C1 conserved region of the methyl-accepting domain was another group specific region,
162 though part of this region was conserved between groups 1,3 and 4 (AAs ~220-260) making this
163 a third region specific to Aer.g2. Finally, the C-terminal region after the last conserved C3 region
164 was specific for each group, especially Aer.g2 as it has an extension that the other groups do not.
165 Thus within the Aer family, aer.g2 is distinguished by its N-terminal region, central region and C-
166 terminal extension while aer.g5 is notable in its complete differentiation the entire length of the
167 protein. The other groups, Aer.g1, Aer.g3 and Aer.g4 can be distinguished from one another based
168 on the composition of the variable regions (AAs ~155-165, ~195-225, ~250-275, ~305-360 and
169 ~395-410).

170 **Distribution of Groups within *Pseudomonas***

171 To determine the prevalence of each Aer homolog group within the *Pseudomonas* genus,
172 the number of homologs from each group that each species possessed was counted and a
173 hierarchically clustered heatmap was generated (Figure 3). Aer-2 and CttP were also counted and
174 included. This analysis showed that all included species, except *P. dentrificans* ATCC13867, had
175 an Aer.g1 homolog. Only species related to *P. aeruginosa* had Aer-2 and CttP, which were always
176 found together. *P.* sp 21, *P. nitroreducens* HBP1, *P. resinovorans* NBRC 106553 and *P.*
177 *pseudoalcaligenes* KF707 were the only species that had Aer-2/CttP and additional Aer homologs,
178 the first two possessing an Aer.g2 homolog, *P. resinovorans* having an Aer.g4 homolog and
179 KF707 having both. Only genes from Aer.g1, Aer.g2 and Aer.g3 were duplicated, and duplications
180 were rarer than possession of homologs from multiple groups. Duplications of Aer.g1 were

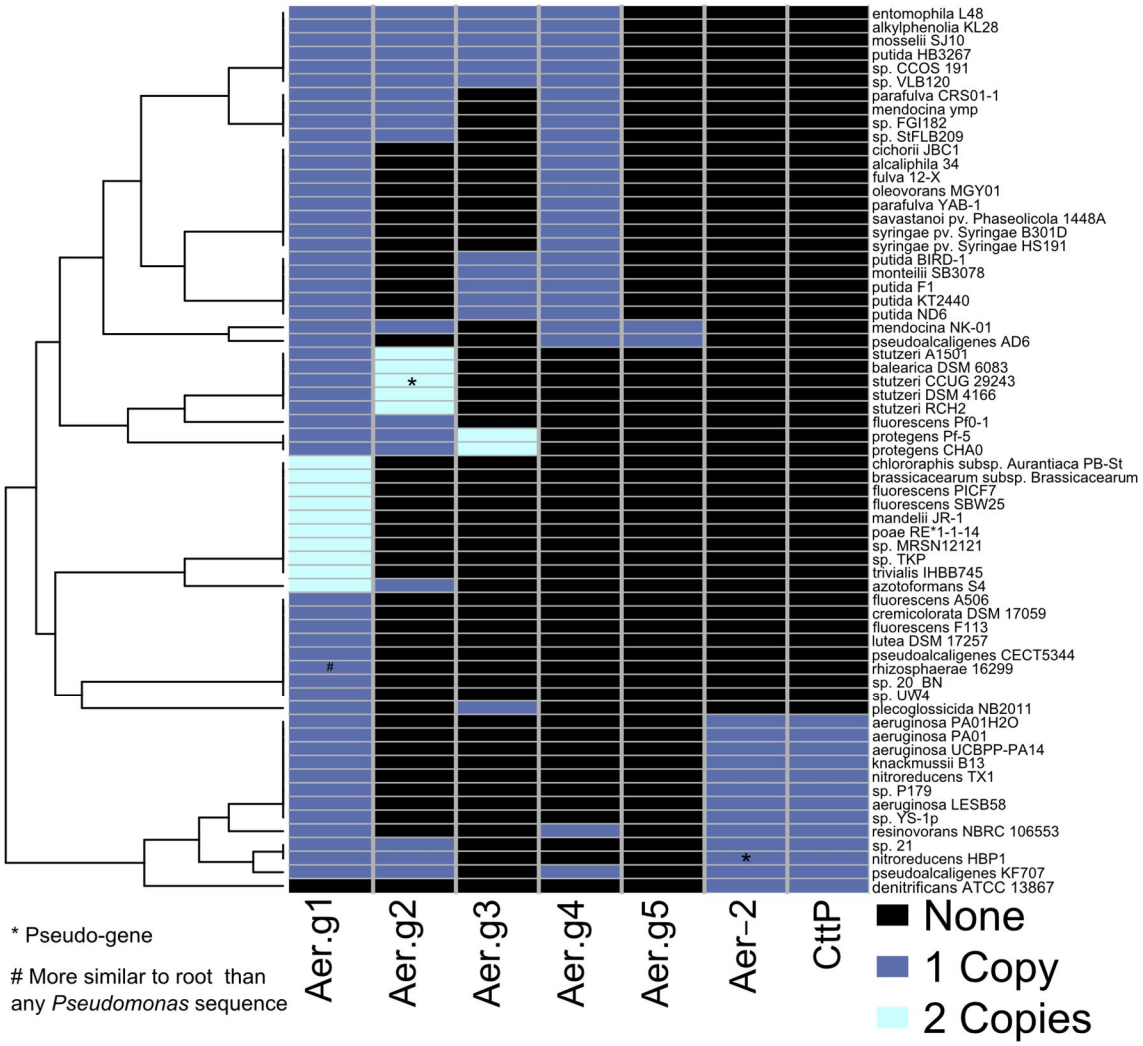

Figure 3: Hierarchically clustered heatmap showing the presence and number of Aer homologs, Aer2 and Cttp in select *Pseudomonas* species. Species were clustered using the Bray-Curtis distance metric and average linkages. The number of homologs that each strain possess is indicated by the box colour: zero (black), one (blue), two (cyan). Pseudogenes are marked by asterisks (*). The homolog of *P. rhizosphearae* is more similar to non-*Pseudomonas* sequences and is marked by a pound (#).

restricted to *P. fluorescens* and related subgroups (1-8) (33). 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.

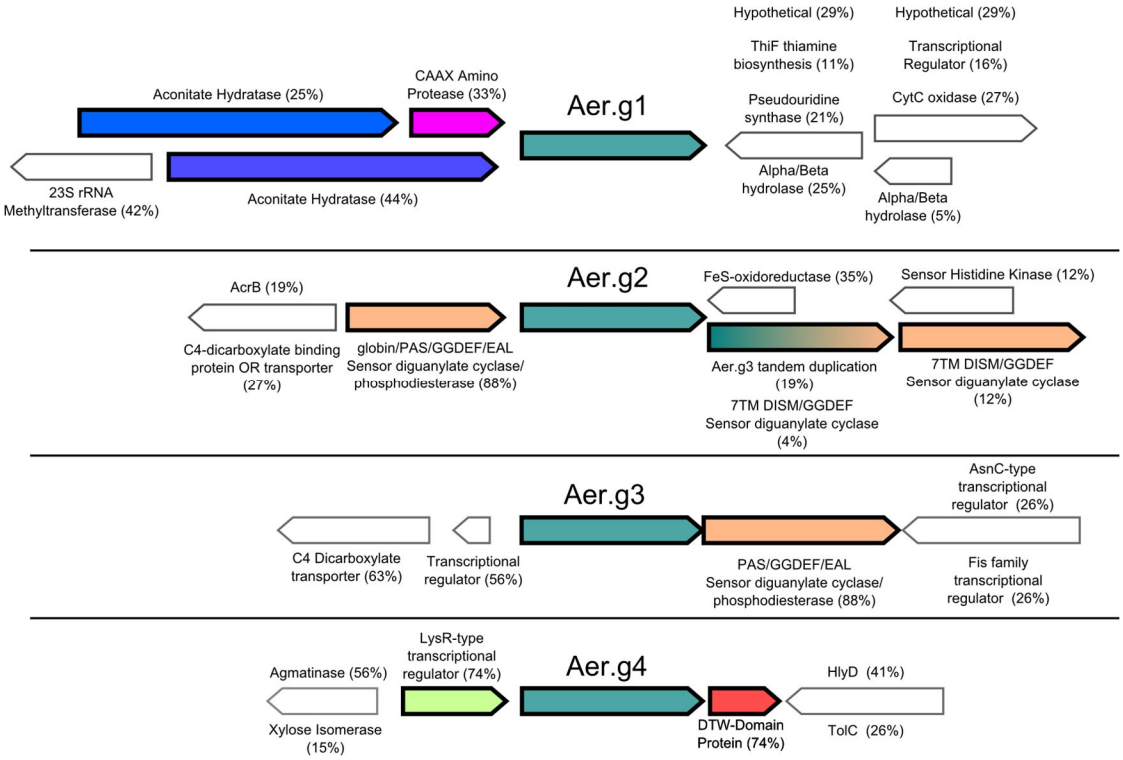

Figure 4: Frequency of occurrence and orientation of genes upstream and downstream from *aer* homologs from select *Pseudomonas* species. Coloured and bolded genes are expected to be co-transcribed based on their orientation and high frequency of occurrence with *aer*. Numbers beside gene functions indicate the frequency that they were found within the 144 sequences examined. Gene lengths are approximate, *aer* is about 1.5kb long.

Overall, the tree clustering *Pseudomonas* species based on possession of Aer homologs roughly matched the known species phylogeny, though there were several noteworthy instances of variability at the strain level. The presence of Aer-2/CttP created a major separation, however this separation matches that of the aeruginosa subgroup. Only 1/2 *P. parafulva* strains had an Aer.g2 homolog and only 2/4 *P. fluorescens* strains had the same Aer homolog profile. 3/4 of *P. putida* strains were the same, though *P. putida* HB3267 had a homolog from all four groups. *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 variation as the third strain

208 included in the analysis, *P. pseudoalcaligenes* CECT5344, only possessed a single Aer.g1
209 homolog.

210 **Evidence for Horizontal Gene Transfer of Aer Homologs**

211 Based on the varied distribution of Aer homologs within the genus *Pseudomonas* it was
212 hypothesized that horizontal gene transfer (HGT) could have caused this. To examine this
213 possibility, the Aer phylogenetic tree (Figure 1) was compared to the known *Pseudomonas* species
214 phylogeny (33). Additionally, upstream and downstream regions of the *aer* genes were examined
215 to determine whether the genomic context was consistent across species (Figure 4) and whether
216 they were associated with mobile elements (Supplementary Table 1). Inspection of the
217 phylogenetic tree revealed 3 sequences that were highly divergent from the expected species
218 phylogeny (Figure 1, green stars). These 3 sequences (WP_015271024.1, *P. putida* HB3267;
219 WP_014754514.1, *P. putida* ND6; and WP_013791017.1, *P. fulva* 12-X) were each found
220 clustered with sequences from unrelated species whereas other Aer sequences were consistently
221 clustered with sequences from closely related species. Of these 3 genes only the *aer.g3* gene from
222 *P. putida* ND6 was associated with mobile elements.

223 The first two upstream and downstream genes of each *aer* gene were identified, as well as any
224 nearby mobile elements (inverted repeats, transposases, integrases) (Supplementary Table 2). The
225 frequency of each of the associated genes and the general genomic context for each Aer homolog
226 were examined to identify the most commonly associated genes (summarized in Figure 4). Each
227 homolog was part of an apparently unique operon: Aer.g1 with an aconitate hydratase (69%) and
228 a CAAX aminoprotease (33%), Aer.g2 with a PAS-containing diguanylate
229 cyclase/phosphodiesterase (88%), Aer.g3 with a different PAS-containing diguanylate
230 cyclase/phosphodiesterase (88%) and Aer.g4 with a LysR-type transcriptional regulator (64%) and

231 a DTW-domain containing protein (74%). Beyond these, the second upstream and downstream
232 genes tended to vary more widely, though some were clearly conserved: Aer.g1, 23S rRNA methyl
233 transferase (42%); Aer.g3, C4 dicarboxylate transporter (63%); Aer.g4, agmatinase (56%). Still
234 this indicates clear variability in the genomic context of homologs within each group. Also of note
235 were two other diguanylate cyclases, occurring at much lower frequencies than those mentioned
236 above (Supplementary_Table_Genomic_Context). As Aer.g5 had only two members and their
237 genomic contexts were identical, they are not presented here.

238 Mobile elements were found within 5Kb of *aer* homolog genes (Supplementary Table 1).
239 Inverted repeats were most common (20%), followed by transposases (4%) and integrases (1%).
240 Aer.g5 never had any mobile elements associated, Aer.g1 was the least common, then Aer.g3, then
241 Aer.g2 and Aer.g4 had the highest frequency of association with mobile elements. When mobile
242 elements were found, they were often located immediately up and/or downstream of the *aer*
243 homolog and its associated gene(s) described in the above section.

244 **Phenotypic Results**

245 To investigate the functions of Aer homologs, CttP and Aer-2, and to see whether they
246 were linked, deletion mutants were constructed in the one species of *Pseudomonas* that possessed
247 3 Aer homologs Aer-2 and CttP: *P. pseudoalcaligenes* KF707. Single and combination deletions
248 were constructed using 2-step homologous recombination, the suicide vector pG19II and SacB
249 sucrose counter-selection (35). Energy-taxis was tested using soft agar swim plates (36)
250 (sometimes called 'swarm' plates, though the motility is definitely swimming not swarming). In
251 this assay, a small amount of cells are inoculated into a plate by stabbing a needle covered in
252 culture into an agar plate containing a low (0.3%) percentage of agar and a high amount of carbon
253 source (50mM) in minimal salts. As the culture grows and consumes the carbon source at the point

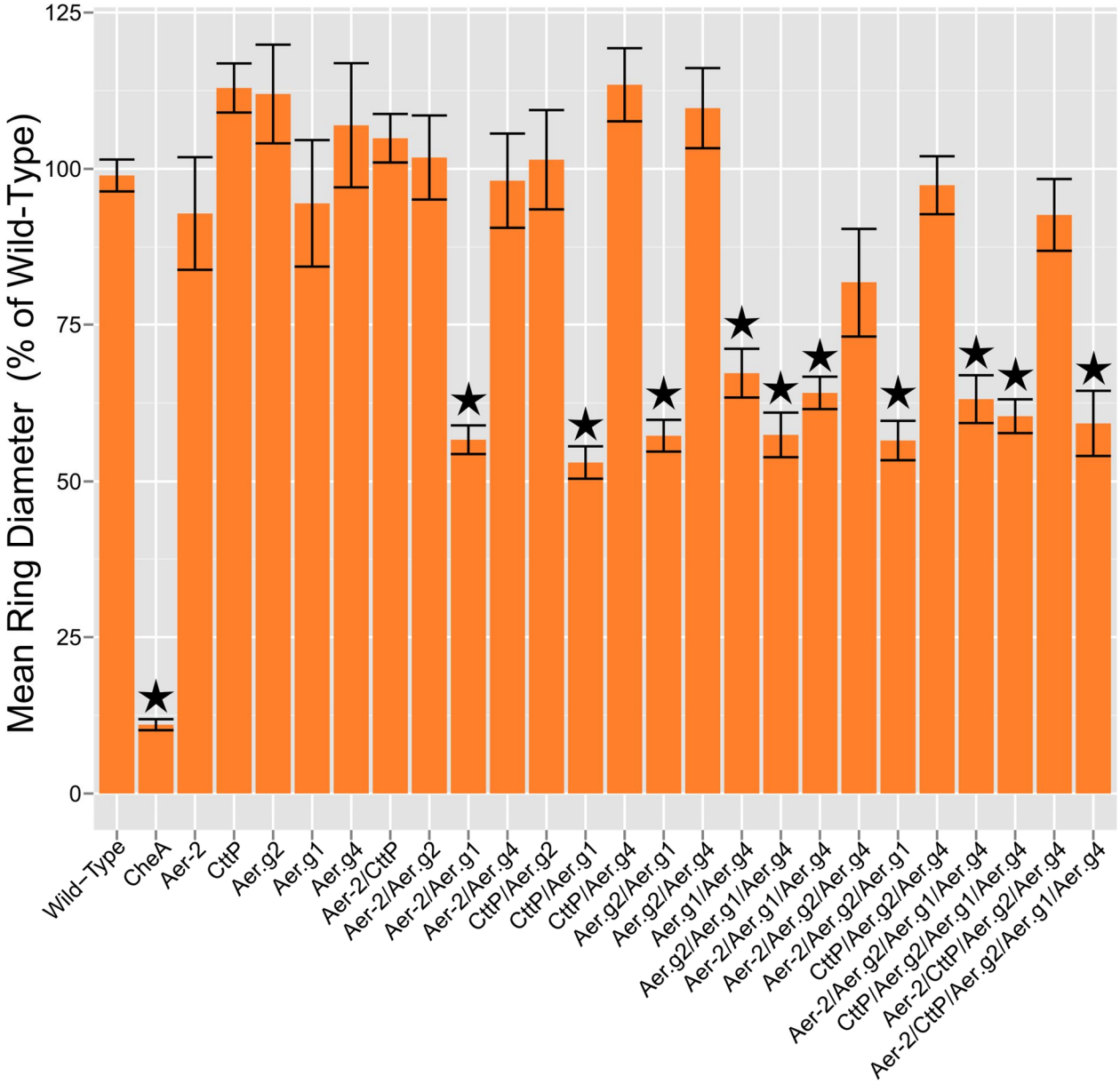

Figure 5: Normalized energy-taxis growth diameters in 50mM pyruvate of strains of *P. pseudoalcaligenes* KF707 with deletions of *aer2, cttP* and *aer* homologs. Bars indicate the average growth diameter, normalized to the wild-type, at both 24h and 48h from at least 3 experimental replicates. Wild-type 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. The *cheA::KmR* mutant was not grown with antibiotic present.

of inoculation it will produce a ring as energy-tactic cells seek out a better place to grow. Non-motile or non-energy tactic cells will produce a smaller ring (Supplementary Figure 4). All 24 mutants generated in this study, along with the wild-type and a CheA::KmR insertional

inactivation non-motile mutant (37) were tested in triplicate in media containing pyruvate (Figure 5) and succinate (Supplementary Figure 5). Ring diameters were normalized to the wild-type, then tested for differences using Tukey's Honest Significant Differences test (Supplementary Table 3).

The diameter of the chemotactic negative CheA mutant was about 10% of the wild-type, whereas energy-taxis negative mutants with diameters about 60% smaller were significantly different from the wild-type (Figure 5). All the singular deletions of Aer.g1, Aer.g2, Aer.g4, Aer-2 and CttP had no effect on energy-taxis; only strains that had Aer.g1 deleted in combination with at least one other receptor were energy-taxis negative. This included both homologs, Aer-2 and CttP. Conversely, deletion of any combination of these receptors, including the quadruple Aer-2/CttP/Aer.g2/Aer.g4 mutant did not adversely affect energy-taxis. The Aer-2/Aer.g2/Aer.g4 triple mutant had a diminished diameter, about 80% of the wild-type, but this was not significant. These results from plates containing pyruvate were similar to those obtained using succinate (Supplementary Figure 5), however those swim diameters were far more variable resulting in larger error bars and lack of significant differences. The observed differences in swim diameters were not due to differences in growth rate as the rate of growth of the swim diameters was similar between all strains (Supplementary Figure 6, Supplementary Table 4). All strains were also tested for chemotaxis towards pyruvate, glucose (Supplementary Figure 7) and succinate (data not shown) using traditional swim plate assays, and all except the CheA::KmR mutant were able to swim towards, pyruvate, succinate and glucose.

Discussion

In *Pseudomonas* the MCP Aer has been characterized as a receptor for the conflated processes of aerotaxis and energy-taxis. The presence of multiple homologs in some species, of

288 which only one is functional, has led to confusion with a putative oxygen receptor as the former is
289 named Aer2 and the latter Aer-2. Here we combined a bioinformatics analysis of Aer within the
290 genus *Pseudomonas* with a molecular biology characterization of Aer homologs in the one species
291 (*P. pseudoalcaligenes* KF707) that possessed 3 Aer homologs, Aer-2 and the putative
292 tetrachloroethylene receptor CttP, revealing that energy-taxis and aerotaxis are linked but distinct
293 processes and that Aer is a family of patchily distributed receptors in *Pseudomonas.*

294 **Sequence and genetics classify Aer homologs into 5 groups**

295 The phylogeny and alignment of Aer sequences implied there were up to 7 different groups,
296 but comparison of sequence differences using SHMR (23) showed there were actually 5. All
297 groups varied in the same regions of the protein, except Aer.g2 which had unique regions at the N
298 and C termini and between the transmembrane helices and conserved methyl-accepting domain.
299 The most conserved domains were unsurprisingly the PAS ligand binding domain and CheW
300 interface region which confer the necessary functions of FAD cofactor binding/signal reception
301 (38) and signal transduction, respectively. Two other conserved regions were identified, on either
302 side of the CheW interface indicating their importance for the core function of the protein. The
303 conserved regions in the signaling region were better conserved than the predicted PAS domain,
304 though these domains are known to differ at the sequence level (38) so it is likely that proteins
305 from the different groups function similarly. Sequence based group assignments were supported
306 by the genomic context of the corresponding genes; there tended to be shared synteny in the
307 immediate vicinity of the *aer* genes from each group, though this also supported the existence of
308 additional groups. Including the branch of Aer.g1 (Figure 1, dotted box) within Aer.g1 was
309 supported by the SHMR analysis however these genes did not share synteny with the rest of Aer.g1.
310 These sequences represent the only duplications of Aer.g1 and only occurred in species which had

311 no other Aer homologs (except *P. azotoformans* S4). As the protein sequences do not differ as
312 much as the other groups, and the genomic context within this group differed, this may represent
313 a recent duplication event within these lineages. Phenotypic characterization of the two Aer
314 receptors in these species could be particularly enlightening to understanding why only some
315 *Pseudomonas* species have extra *aer* genes.

316 **Aer.g1 is ancestral, prevalent and functional**

317 The Aer.g1 homolog group was the closest to the root, with the exception of Aer from *P.*
318 *rhizosphearae*, indicating this group is ancestral to the other Aer groups. It was also the most
319 prevalent as all the *Pseudomonas* species included in this study had at least one homolog from this
320 group except *P. denitrificans* ATCC13867. This means that Aer is not part of the *Pseudomonas*
321 core genome, though it is likely ancestral and was lost in *P. denitrificans*. Interestingly this also
322 implies that the oxygen receptor Aer-2 is sufficient for the lifestyle of this organism, named for its
323 ability to use a terminal electron acceptor other than oxygen. All of the previously characterized
324 functional Aer homologs, from *P. aeruginosa* PA01 (8), *P. putida* F1 (39) and *P. putida* KT2440
325 (9) are part of the homolog group Aer.g1, which was demonstrated to be pivotal in *P.*
326 *pseudoalcaligenes* KF707. The functional Aer homolog from *P. putida* PRS2000 (11) was not
327 included in this study as this species genome has not been fully sequenced. The receptors from *P.*
328 *putida* F1 and KT2440 were both named Aer2 but the phylogeny presented here indicates they
329 belong to the Aer.g1 homolog group. As KF707 is the only strain where the other Aer homologs
330 have been shown to have semi-redundant functionality with Aer.g1, it is safe to predict that the
331 Aer.g1 homolog from other species will be the receptor responsible for energy-taxis.

332 Almost all *aer.g1* genes were associated with an aconitase gene, indicating that they are
333 likely co-transcribed. In *Helicobacter pylori,* TlpD, which controls tactic behaviour in low-energy

conditions has been found to interact with aconitase (40). Our results thus indicate that association with aconitase is a common feature of energy-taxis receptors. While Aer cannot obtain electrons for reducing the FAD co-factors from aconitase as the isomerization of citrate to isocitrate is a non-redox process, this suggests that isocitrate dehydrogenase may be the step of the citric acid cycle that is being sensed for energy-taxis.

A gene encoding a predicted CAAX amino protease was found adjacent to one third of Aer.g1 homologs. This type of membrane-bound protease aids in prenylating proteins to ease their membrane localization (41). The associated Aer.g1 proteins did not have the expected CAAX motif though they had high sequence similarity to each other. There were no obvious features at the sequence level that distinguished these sequences completely from other Aer homologs indicating that the genetically associated CAAX protease may not necessarily be involved in their functioning.

Other Aer groups likely have unique functions

As the 5 characterized Aer.g1 homologs appear to be broad-purpose energy-taxis sensors, this implies that the other homolog groups may have a related, but different or more specific function. The three Aer homologs of *P. putida* KT2440 were found to be differentially expressed (9), supporting this hypothesis. In *E. coli* Aer is also a thermosensor (42) indicating another possible role or they could aid in tuning the energy-taxis response similarly to how Tsr and Tar from *E. coli* enable taxis towards the ideal pH from both higher and lower initial pHs (43). The function of the additional Aer homologs could also be related to biofilm formation as Aer.g2 and Aer.g3 were associated with their own PAS-domain containing diguanylate cyclase/phosphodiesterase. Deletion of the diguanylate cyclase/phosphodiesterase associated with Aer.g3 in *P. putida* KT2440 caused a general defect in motility (9) and was later shown to be a bi-

functional cyclic di-GMP phosphodiesterase and diguanylate cyclase formation (44). These two homolog groups may thus be involved in energy-sensing behaviour in biofilms as c-di-GMP is an important regulator in the transition from planktonic to sessile growth modes, though it will be interesting to see if the putative diguanylate cyclase/phosphodiesterase associated with Aer.g2 is also bifunctional and whether they both respond to the same signal. Aer.g4 could also have its own distinct function, possibly transducing its signal not only to the *che* pathway but also to the transcriptional regulator adjacent to it allowing for gene regulation in response to energy-sensing. The function of Aer.g5 may be even more distant as it is the least similar from the others, though it is likely a specific niche function since only two of the included species possessed copies. This receptor definitely evolved in *Pseudomonas* though as all its closest relatives (by BLAST) are *Pseudomonas* proteins. These results thus raise many questions about energy sensing and how it is integrated into other cellular responses than flagellar swimming.

aer has been horizontally transferred

Multiple lines of evidence indicate that genes from the *aer* group have been horizontally transferred. Foremost is the observed variation at the strain level as there were several noted examples of differences in the number of homologs possessed by strains of the same species. 3 specific sequences, from *P. putida* HB3267, ND6 and *P. fulva* 12-X clustered beside Aer sequences from unrelated species indicating they may have been obtained from those species. The observation that many strains did not have the same complement of Aer homologs also supports this notion. For example, of the 3 *P. pseudoalcaligenes* strains examined here, each one had a drastically different complement of genes (CECT5344, *aer.g1*; AD6, *aer.g1/aer.g4/aer.g5*; KF707, *aer.g1/aer.g2/aer.g4/aer-2/cttP*). This demonstrates the ancestrality of *aer.g1* and the variability of other members, but also makes KF707 an outlier from the species. It having *aer-2* and *cttP*

380 indicates either that these have also been transferred horizontally or KF707 is more closely related
381 to *P. aeruginosa* than the other strains. As only the aeruginosa sub-group has those two genes they
382 were also likely horizontally transferred, albeit less recently than the various *aer* homologs.

383 The observed variation in the distribution of Aer homologs is unsurprising as even within
384 the *P. aeruginosa* pan-genome from 7 strains there are 2,000 accessory genes compared to 5,000
385 core (45). The fact that *aer.g1* almost qualifies as part of the genus core genome (*P. dentrificans*
386 being the only outlier) indicates that is a very useful gene in the varied environments inhabited by
387 *Pseudomonas*. The existence of the various homologs speaks to their utility as they each may
388 confer a more specific function. As many *Pseudomonas* species live in the soil and rhizosphere,
389 which promotes horizontal gene transfer (46), there would be ample opportunity for these *aer*
390 homolog genes to spread. Interestingly they appear to be restricted to *Pseudomonas* still as reverse
391 BLAST searches consistently turned up *Pseudomonas* sequences as the closest hits. Even more
392 compelling are the associated genes which were transferred alongside. The bifunctional
393 diguanylate cyclase/phosphodiesterase adjacent to *aer.g3* in *P. putida* KT2440 influences motility
394 (9) but this is only 1 of 4 c-di-GMP processing gene products found near *aer* homologs. The
395 function of these genes and the transcriptional regulator and DTW-domain protein beside *aer.g4*
396 are unknown and likely represent interesting connections of energy-sensing to other cellular
397 processes.

398 **Aer-2 and CttP contribute to energy-taxis in *P. pseudoaligenes* KF707**

399 CttP was characterized as a receptor for positive chemotaxis towards chlorinated ethylenes
400 (13), but this function may be fortuitous similar to how Tar and Tsr from *E. coli* can mediate
401 attractant and repellent responses to phenol through its direct interaction with the transmembrane
402 helices or cytoplasmic HAMP signaling domains (47). Here we showed that co-deletion of *cttP*

403 and a*er.g1* in *P. pseudoalcaligenes* KF707 abolished energy-taxis indicating that it is also an
404 energy-taxis receptor. This is likely untrue though as prediction of the domain architecture of CttP
405 indicates that it begins with two-transmembrane helices and does not have an N-terminal ligand
406 binding domain like other MCPs. Instead it has the unusual feature of a long (>100 AA) C-terminal
407 extension after the CheW interface region. It could thus be possible that CttP does not directly
408 sense any signal, but assists in forming MCP receptor complexes and/or linking them with the Che
409 complex. Overexpression of CttP (then called McpA) abolished all chemotaxis in both *P.*
410 *aeruginosa* and *E. coli* (36) further implicating it as general chemotaxis protein. MCPs have been
411 proposed to amplify signals by propagating ligand-binding induced changes to adjacent MCPS
412 (48). CttP could be performing such a function as it is expressed in stationary phase (49) and
413 overexpression induces a smooth-swimming phenotype (36).

414 The same abolishment of chemotaxis was observed when Aer-2 (McpB) was
415 overexpressed in both organisms (36). This could be related to its function of being necessary for
416 the complexation of the other Che proteins located in the same operon *che2* (49). Along with this
417 necessity, Aer-2 has a C-terminal extension that specifically allows CheR2, and only CheR2, to
418 methylate it (50), making it appear that the entire Che2 pathway exists to transduce the signal
419 sensed by Aer-2. In the current study we found that deletion of either *cttP* (*mcpA*) or *ae-r2* (*mcpB*)
420 in combination with *aer.g1* abolished energy-taxis indicating that there is a connection between
421 the receptors from the *che2* operon and energy-taxis. As these receptors are only present in a
422 narrow subsection of *Pseudomonas* species (the aeruginosa subgroup), which varied widely in
423 their number of Aer homologs, it is unlikely that this is a direct connection. The *in vitro*
424 biochemical studies of Aer-2 indicating that it binds oxygen and the demonstration that its
425 expression in *E. coli* causes a repellent response to oxygen (10), along with the original

426 characterization of Aer-2 as an aerotaxis receptor (8), strongly indicate that it is a true aerotaxis
427 receptor that binds oxygen. This fits with the observations we made in *P. pseudoalcaligenes*
428 KF707 where the deletion of *aer.g1* and *aer-2* resulted in an energy-tactic negative phenotype. In
429 the soft agar energy-taxis swim plate assay used in this study, oxygen as well as metabolizable
430 carbon source would be diminished at the center of the growth ring. This would allow the energy-
431 sensing of Aer and oxygen sensing of Aer-2 to work together, e.g. in the Aer.g1 and Aer.g2/Aer.g4
432 mutants that did not display diminished energy-taxis.

433

434 **Conclusions**

435 In *Pseudomonas* Aer does not simply refer to a single receptor for energy-taxis, but to a
436 family of receptors of mostly unknown function. The most ancestral Aer, here named Aer.g1, is
437 indeed an energy-taxis receptor, though other proteins may contribute to its function, including
438 other Aer family members. In *P. pseudoalcaligenes* KF707, unique among *Pseudomonas* species
439 for having 3 Aer homologs, Aer-2 and CttP, also uses these other receptors for energy-taxis.
440 Though as most species do not have these receptors and other evidence indicates their true purpose
441 this may be a minor function. Apart from Aer.g1, Aer family members, Aer-2 and CttP have likely
442 been horizontally transferred within the genus indicating that all these receptors play important
443 roles in *Pseudomonas* physiology.

444

445

Materials and Methods

Protein Sequences

All sequences were obtained using NCBI databases and tools (14) in May 2016. BLAST searches (15) (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 true Aer homologs. Expect values were always below $1x10^{-100}$. Two 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. Two more sequences were removed for similar reasons. Sequence accession numbers were thus obtained from the international nucleotide sequence database collaboration (INSDC) (16). NCBI Entrez was used to obtain FASTA formatted sequences which were then compiled into a single file. A sequence to be used as a root was also acquired by BLASTing NP_250252.1 against the NCBI database, but excluding any *Pseudomonas* sequences from the results. WP_0027981445.1 from an unknown gamma proteobacterium L18 was thus obtained.

Alignment

Two alignments were made, with and without the root, using COBALT (17). The default settings were used as adjustment did not improve the alignment noticeably. As this alignment had numerous unnecessary gaps introduced, these were manually removed by adjusting the alignment in Jalview (18), which was then re-aligned using MUSCLE (19). Names were cleaned up using a custom script in R (20). The alignments were input into the Mobyle web platform (21) and maximum likelihood phylogenies were constructed using PHYML (22). PHYML options: Amino-

469 Acids, 100 bootstrap replicates, JTT amino-acid substitution model, BEST tree topology search
470 operation, tree topology, branch length and rate parameters were optimized. Consense was then
471 used to generate a consensus tree from the bootstrap replicates using the majority rule (extended),
472 and only treating the tree as rooted when the non-pseudomonas root was included.

473 Sequence Harmony and Multi-Relief Analysis to Determine Groups

474 The multi-Harmony server was used to apply sequence harmony and multi-relief (SHMR)
475 to validate groupings made based on the ML tree and alignment. Groups were manually decided
476 based on the tree topology and corresponding alignment, initially making for 7 groups. These
477 seven groups were compared in pair-wise fashion using SHMR (23). Thus for each pair of groups
478 a score was calculated for each amino acid that indicated how conserved it was within groups and
479 how divergent it was between groups, a score of 1 indicating perfect conservation within and
480 perfect divergence between. Empirical cumulative distribution functions (ECDF) were plotted for
481 each comparison and the percent of AAs above the 0.8 cutoff recommended by the SHMR authors
482 was determined. As most comparisons resulted in 25% of AAs reaching this threshold, the two
483 pairs of groups that, when compared, only had 10% of AAs above the cutoff, were deemed
484 incorrect group assignments. These groups were merged and the process repeated to produce 5
485 groups that all had ~25% of AAs above the cutoff threshold. Group 5 was consistently excluded
486 as it only contains 2 highly divergent sequences.

487 Comparison of Groups

488 Unique regions of each group of Aer homologs were identified by comparing the SHMR
489 scores for each pair-wise comparison with the overall conservation score of each AA and the
490 domain architecture of the Aer protein. Conservation scores were obtained from Jalview (18), and
491 along with the SHMR scores were smoothed and plotted using ggplot2 (24) in R. Smoothing was

performed by calculating the average of the 3 proceeding and following AAs for each position. The domain architecture of Aer was obtained from the conserved domain database (CDD) (25).

Distribution of Groups

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

Detection of Evidence of Horizontal Gene Transfer

Graphical representations of the Aer homologs nucleotide sequences were manually inspected on NCBI. The first two genes upstream and downstream of the *aer* homolog were noted, along with any mobile elements (transposase, integrase and inverted repeats) within 5kb. For each different upstream and downstream gene their frequency of occurrence was calculated for each homolog group. The frequency of occurrence of each type of mobile element was also calculated, as well as for the complete set of homologs.

Culture Growth

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

514 Generation of Deletion Constructs and Mutants

515 Nucleotide sequences for Aer.g1, Aer.g2, Aer.g4, Aer-2 and CttP were obtained from the
516 draft genome sequence of *P. pseudoalcaligenes* KF707 (26). Primers for a ~500bp region up and
517 downstream of each region were generated using Primer BLAST (27). Benchling was used to alter
518 the primers, adding BamHI or HinDIII restriction sites to the outer primers and the reverse-
519 complement of the other inner primer to each of the internal primers. Genomic DNA was isolated
520 by the phenol/chloroform method (28). In separate PCR reactions the upstream and downstream
521 fragments were amplified using Hi-Fidelity (HF) Enzyme mix (Fisher Scientific, USA). Fragments
522 were purified by gel extraction using an EZDNA kit (Omega Bio-Tek, USA) then pooled and used
523 as the template for the second PCR reaction using only the outer primers. The pG19II vector(29),
524 purified using an EZDNA plasmid mini kit II (Omega Bio-Tek, USA), and insert were digested
525 using BamHI and HinDIII (Invitrogen, USA). Digestion products were purified then ligated
526 together using T4 ligase (Invitrogen, USA. Ligations were transformed either directly into *E. coli*
527 Top10F' or first into DH5α chemical competent cells using standard methods (28). White colonies
528 were picked from LB X-Gal gentamycin (20µg/mL) plates and their plasmids isolated and
529 screened for a ~1kb bandshift. Those with the appropriate shift were sequenced (Eurofins, USA)
530 to confirm the correct insert sequence. To delete the genes from the *P. pseudoalcaligenes* KF707
531 genome, the deletion construct containing plasmids were introduced by conjugation. Cultures of
532 *E. coli* HB101 carrying the helper plasmid pRK2013 (30) and *E. coli* Top10F' carrying the deletion
533 construct in pG19II were grown to early log-phase (OD ~0.3) along with the KF707 wild-type, or
534 later, deletion mutants. Donor, helper and recipients were mixed and plated, grown overnight then
535 the cell mass was collected and spread on AB glucose plates (5g/L glucose, minimal salts media)
536 containing 20µg/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 + 20µg/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.

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 24 and 48h either manually using a ruler or digitally using a photograph and ImageJ (31). The experiment was repeated at least 3 times for all strains in each media, always including at least one wildtype to be used as a normalizing control. Collected data were processed in R to normalize the size of the growth diameter to the corresponding wildtype size at 24 and 48h. The data were analyzed in a number of ways and it was determined that pooling the normalized data from 24 and 48h 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 (20) 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.

560 Chemotaxis Swim Plates

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

568

569 **Funding Information**

570 This project was funded by Natural Sciences and Engineering Research Council (NSERC) of
571 Canada Discovery Grant RGPIN/219895-2010 to RJT.

572

573 **Acknowledgements**

574 The authors would like to thank Iain George and Dr. Gordon Chua for use of their imaging setup
575 for photographing the swim plates. We thank Dr. Martina Cappelletti for assistance with the
576 colony PCR and her insights into KF707 physiology and genomic interpretation.

577 **References**

578 1. **Wadhams GH**, **Armitage JP**. 2004. Making sense of it all: Bacterial chemotaxis.

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

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

583 4. **Beijerinck MW**. 1893. Ueber atmungsfiguren beweglicher bakterien. Zentrabl Bakteriol
584 Parasitenkd **14**:827–845.

585 5. **Engelmann TW**. 1881. Neue Methode zur Untersuchung der Sauerstoffausscheidung
586 pflanzlicher und thierischer Organismen. Pflüger, Arch für die Gesammte Physiol des
587 Menschen und der Thiere **25**:285–292.

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

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

594 8. **Hong CS**, **Shitashiro M**, **Kuroda A**, **Ikeda T**, **Takiguchi N**, **Ohtake H**, **Kato J**. 2004.
595 Chemotaxis proteins and transducers for aerotaxis in Pseudomonas aeruginosa. FEMS
596 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. **Watts KJ**, **Taylor BL**, **Johnson MS**. 2011. PAS/poly-HAMP signalling in Aer-2, a soluble haem-based sensor. Mol Microbiol **79**:686–99.

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

12. **Parales RE**, **Luu R a.**, **Chen GY**, **Liu X**, **Wu V**, **Lin P**, **Hughes JG**, **Nesteryuk V**, **Parales J V.**, **Ditty JL**. 2013. Pseudomonas putida F1 has multiple chemoreceptors with overlapping specificity for organic acids. Microbiol (United Kingdom) **159**:1086–1096.

13. **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.

14. **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.

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

16. **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.

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

620 18. **Waterhouse AM**, **Procter JB**, **Martin DMA**, **Clamp M**, **Barton GJ**. 2009. Jalview
621 Version 2--a multiple sequence alignment editor and analysis workbench. Bioinformatics
622 **25**:1189–91.

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

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

627 21. **Neron B**, **Menager H**, **Maufrais C**, **Joly N**, **Maupetit J**, **Letort S**, **Carrere S**, **Tuffery**
628 **P**, **Letondal C**. 2009. Mobyle: a new full web bioinformatics framework. Bioinformatics
629 **25**:3005–3011.

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

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

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

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

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

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

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

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

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

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

655 32. **Bodilis J**, **Nsigue Meilo S**, **Cornelis P**, **De Vos P**, **Barray S**. 2011. A long-branch
656 attraction artifact reveals an adaptive radiation in pseudomonas. Mol Biol Evol **28**:2723–
657 6.

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

660 34. **Livermore JA**, **Emrich SJ**, **Tan J**, **Jones SE**. 2014. Freshwater bacterial lifestyles
661 inferred from comparative genomics. Environ Microbiol **16**:746–758.

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

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

669 37. **Tremaroli V**, **Fedi S**, **Tamburini S**, **Viti C**, **Tatti E**, **Ceri H**, **Turner RJ**, **Zannoni D**.
670 2011. A histidine-kinase cheA gene of Pseudomonas pseudoalcaligens KF707 not only
671 has a key role in chemotaxis but also affects biofilm formation and cell metabolism.
672 Biofouling **27**:33–46.

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

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

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

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

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

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

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

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

693 46. **Sengeløv G**, **Kristensen KJ**, **Sørensen AH**, **Kroer N**, **Sørensen SJ**. 2001. Effect of
694 Genomic Location on Horizontal Transfer of a Recombinant Gene Cassette Between
695 Pseudomonas Strains in the Rhizosphere and Spermosphere of Barley Seedlings. Curr
696 Microbiol **42**:160–167.

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

699 48. **Wolanin PM**, **Stock JB**. 2004. Bacterial Chemosensing: Cooperative Molecular Logic.
700 Curr Biol **14**:R486–R487.

701 49. **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.

50. **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.

Supplementary Figure and Table Captions

Supplementary Figure 1: Maximum likelihood consensus tree showing phylogenetic relationship between Aer homolog protein sequences from select *Pseudomonas* species, rooted using the closest available non-*Pseudomonas* sequence. Tree was generated unrooted, then then rooted to the known outgroup. Numbers at nodes indicate bootstrap support values from 100 replicates.

Supplementary Figure 2: Empirical cumulative distribution function of amino acid multi-relief scores for pair-wise comparisons of preliminary Aer homolog groups. Only scores above 0.5 are shown, and accumulation above 40%. The dotted line at 0.8 allows to easily see what percent of amino acid positions for each comparison are above this cutoff. E.g. only 10% of amino acid positions in the 1vs2 comparison and even less in the 3Avs3B comparison are above this cutoff. Conversely, there is between 25% and 35% of amino acid positions above the cutoff for all other comparisons.

722 Supplementary Figure 3: Empirical cumulative distribution function of amino acid multi-relief
723 scores for pair-wise comparisons of adjusted Aer homolog groups. Only scores above 0.5 are
724 shown, and accumulation above 40%. The dotted line at 0.8 allows to easily see what percent of
725 amino acid positions for each comparison are above this cutoff. All comparisons show that
726 between 25% and 25% of amino acid positions are above the cutoff.

727

728 Supplementary Figure 4: Photographs of energy-taxis swim plates after 24h of growth at 30°C in
729 50mM pyruvate. Energy-taxis swim plates of *P. pseudoalcaligenes* strains with deletions in *aer-2*,
730 *cttP* and *aer* homologs after 24h growth at 30°C in 50mM pyruvate.

731

732 Supplementary Figure 5: Normalized energy-taxis growth diameters in 50mM succinate of strains
733 of *P. pseudoalcaligenes* KF707 with deletions of *aer-2, cttP* and *aer* homologs. Bars indicate the
734 average growth diameter, normalized to the wild-type, at both 24h and 48h from at least 3
735 experimental replicates. Wild-type strains were normalized to the mean of technical replicates
736 within each experiment. Error bars indicate standard error. Stars indicate significant differences
737 from the wild-type based on Tukey's Honest Significant Differences test with a confidence value
738 of 0.95. The *cheA::KmR* mutant was not grown with antibiotic present.

739

740 Supplementary Figure 6: Mean raw and normalized growth rates of energy-taxis swim-growth-
741 diameters of strains of *P. pseudoalcaligenes* KF707 with deletions of *aer-2, cttP* and *aer* homologs.
742 Growth rates were calculated by dividing the diameter at 48h by the diameter at 24h. Normalized

743 rates were calculated from diameters that were normalized to the wild-type for each experiment at
744 each time point. Stars indicate significant differences from the corresponding wild-type according
745 to Tukey's Honest Significant Differences test.

746

747 Supplementary Figure 7: Chemotaxis swim assays of *P. pseudoalcaligenes* strains with deletions
748 in *aer-2, cttP* and *aer* homologs. Strains were grown overnight, concentrated then spotted on
749 minimal salts plates containing 0.3% agar. Either 50mM pyruvate in 1.5% agar or crystals of
750 glucose were placed in the centre of the plates. Photographs were taken after 24h.

751

752 Supplementary Table 1: Frequency of association of mobile elements within 5kb of *aer* genes.

753

754 Supplementary Table 2: Frequency of occurrence of genes neighbouring *aer* homologs in
755 *Pseudomonas*.

756

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

761

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