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

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

**Introduction**

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

Chemotaxis is generally used as a catchall to describe any kind of directed bacterial motion towards or away from a chemical, but for the purposes of this study chemotaxis, aerotaxis and energy-taxis must be more strictly defined. Chemotaxis refers to directed swimming mediated by binding of a ligand either directly or through a periplasmic binding protein to an MCP. Distinguishing between aerotaxis and energy-taxis is more important as these have been traditionally conflated, energy-taxis being a relatively new term. Here aerotaxis will be defined as directed swimming towards oxygen through direct sensing of O₂, and energy-taxis as directed swimming towards a more favourable environment through indirect sensing of energy production within the cell. Aerotaxis was among the first bacterial behaviours ever described as both 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
experimental evidence characterizing CttP as a TCE receptor is clear (13), yet it seems unlikely that this is the natural function of the protein. Due to its appearance as the first gene in the che2 operon, which also contains mcpB (aer-2) we hypothesized that both may have a role in energy-taxis. To investigate these questions, Aer sequences were obtained from 65 Pseudomonas species as well as the presence/absence of Aer-2 and CttP. A phylogeny of Aer was built providing insight into the distribution in the genus and allowing the ‘Aer’ family to be defined. Only one species, P. pseudoalcaligenes KF707 had the surprising feature of possessing 3 Aer homologs, Aer-2 and CttP making it the ideal candidate to investigate if these receptors had related functions. Deletion mutants were generated revealing that in KF707, all 5 of these genes contributed to energy-taxis, though the ancestral Aer homolog was pivotal. Together these results provide a definition of the Aer energy-taxis receptor as a family with varied distribution in Pseudomonas and implicate Aer-2 and CttP in this behaviour.

Results

65 Pseudomonas species were selected for this analysis. Species with completely sequenced genomes were selected, but for those with many strains only a few representatives that have been highly studied were included. Species with incomplete (draft) sequences were also included in an attempt to ensure representation from all major Pseudomonas clades, based on works by (32, 33). 144 protein sequences were obtained from the NCBI database using the P. aeruginosa PA01 Aer sequence (NP_250252.1) as a BLAST (15) query sequence. All hits with >95% sequence coverage were included. These sequences were aligned using COBALT (17) as this alignment algorithm ensures that conserved domains are aligned despite a lack of similarity 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 closest 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. rhizosphaerae* 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 preceding 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.
CheW interface region, each separated by an unconserved, group specific region. The predicted transmembrane domains (AAs ~130-200) were the least conserved region and were not conserved within groups except for the initial portion of these domains (AA ~150). Between this region and the C1 conserved region of the methyl-accepting domain was another group specific region, though part of this region was conserved between groups 1,3 and 4 (AAs ~220-260) making this a third region specific to Aer.g2. Finally, the C-terminal region after the last conserved C3 region was specific for each group, especially Aer.g2 as it has an extension that the other groups do not. Thus within the Aer family, aer.g2 is distinguished by its N-terminal region, central region and C-terminal extension while aer.g5 is notable in its complete differentiation the entire length of the protein. The other groups, Aer.g1, Aer.g3 and Aer.g4 can be distinguished from one another based on the composition of the variable regions (AAs ~155-165, ~195-225, ~250-275, ~305-360 and ~395-410).

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 (Figure 3). Aer-2 and CttP were also counted and included. This analysis showed that all included species, except P. dentrificans ATCC13867, had an Aer.g1 homolog. Only species related to P. aeruginosa had Aer-2 and CttP, which were always found together. P. sp 21, P. nitroreducens HBP1, P. resinovorans NBRC 106553 and P. pseudoalcaligenes KF707 were the only species that had Aer-2/CttP and additional Aer homologs, the first two possessing an Aer.g2 homolog, P. resinovorans having an Aer.g4 homolog and KF707 having both. Only genes from Aer.g1, Aer.g2 and Aer.g3 were duplicated, and duplications 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 Ctp 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. rhizosphaerae* 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. stutzeri* 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/Ctp 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...
Evidence for Horizontal Gene Transfer of Aer Homologs

Based on the varied distribution of Aer homologs within the genus *Pseudomonas* it was hypothesized that horizontal gene transfer (HGT) could have caused this. To examine this possibility, the Aer phylogenetic tree (Figure 1) was compared to the known *Pseudomonas* species phylogeny (33). Additionally, upstream and downstream regions of the aer genes were examined to determine whether the genomic context was consistent across species (Figure 4) and whether they were associated with mobile elements (Supplementary Table 1). Inspection of the phylogenetic tree revealed 3 sequences that were highly divergent from the expected species phylogeny (Figure 1, green stars). These 3 sequences (WP_015271024.1, *P. putida* HB3267; WP_014754514.1, *P. putida* ND6; and WP_013791017.1, *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.

The first two upstream and downstream genes of each aer gene were identified, as well as any nearby mobile elements (inverted repeats, transposases, integrases) (Supplementary Table 2). The frequency of each of the associated genes and the general genomic context for each Aer homolog were examined to identify the most commonly associated genes (summarized in Figure 4). Each homolog was part of an apparently unique operon: Aer.g1 with an aconitate hydratase (69%) and a CAAX aminoprotease (33%), Aer.g2 with a PAS-containing diguanylate cyclase/phosphodiesterase (88%), Aer.g3 with a different PAS-containing diguanylate cyclase/phosphodiesterase (88%) and Aer.g4 with a LysR-type transcriptional regulator (64%) and...
a DTW-domain containing protein (74%). Beyond these, the second upstream and downstream genes tended to vary more widely, though some were clearly conserved: Aer.g1, 23S rRNA methyl transferase (42%); Aer.g3, C4 dicarboxylate transporter (63%); Aer.g4, agmatinase (56%). Still this indicates clear variability in the genomic context of homologs within each group. Also of note were two other diguanylate cyclases, occurring at much lower frequencies than those mentioned above (Supplementary_Table_Genomic_Context). As Aer.g5 had only two members and their genomic contexts were identical, they are not presented here.

Mobile elements were found within 5Kb of aer homolog genes (Supplementary Table 1). Inverted repeats were most common (20%), followed by transposases (4%) and integrases (1%). Aer.g5 never had any mobile elements associated, Aer.g1 was the least common, 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.

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 3 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 (35). Energy-taxis was tested using soft agar swim plates (36) (sometimes called ‘swarm’ plates, though the motility is definitely swimming not swarming). In this assay, a small amount of cells are inoculated into a plate by stabbing a needle covered in 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 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

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

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

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

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

Almost all aer.g1 genes were associated with an aconitase gene, indicating that they are 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*
indicates either that these have also been transferred horizontally or KF707 is more closely related to *P. aeruginosa* than the other strains. As only the aeruginosa sub-group has those two genes they were also likely horizontally transferred, albeit less recently than the various *aer* homologs.

The observed variation in the distribution of Aer homologs is unsurprising as even within the *P. aeruginosa* pan-genome from 7 strains there are 2,000 accessory genes compared to 5,000 core (45). The fact that *aer.g1* almost qualifies as part of the genus core genome (*P. dentrificans* being the only outlier) indicates that is a very useful gene in the varied environments 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 (46), there would be ample opportunity for these *aer* homolog genes to spread. Interestingly they appear to be restricted to *Pseudomonas* still as reverse BLAST searches consistently turned up *Pseudomonas* sequences as the closest hits. Even more compelling are the associated genes which were transferred alongside. The bifunctional diguanylate cyclase/phosphodiesterase adjacent to *aer.g3* in *P. putida* KT2440 influences motility (9) but this is only 1 of 4 c-di-GMP processing gene products found near *aer* homologs. The function of these genes and the transcriptional regulator and DTW-domain protein beside *aer.g4* are unknown and likely represent interesting connections of energy-sensing to other cellular processes.

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

CttP was characterized as a receptor for positive chemotaxis towards chlorinated ethylenes (13), but this function may be fortuitous similar to how Tar and Tsr from *E. coli* can mediate attractant and repellent responses to phenol through its direct interaction with the transmembrane helices or cytoplasmic HAMP signaling domains (47). Here we showed that co-deletion of *cttP*
and aer.g1 in *P. pseudoalcaligenes* KF707 abolished energy-taxis indicating that it is also an energy-taxis receptor. This is likely untrue though as prediction of the domain architecture of CttP indicates that it begins with two-transmembrane helices and does not have an N-terminal ligand binding domain like other MCPs. Instead it has the unusual feature of a long (>100 AA) C-terminal extension after the CheW interface region. It could thus be possible that CttP does not directly sense any signal, but assists in forming MCP receptor complexes and/or linking them with the Che complex. Overexpression of CttP (then called McpA) abolished all chemotaxis in both *P. aeruginosa* and *E. coli* (36) further implicating it as general chemotaxis protein. MCPs have been proposed to amplify signals by propagating ligand-binding induced changes to adjacent MCPS (48). CttP could be performing such a function as it is expressed in stationary phase (49) and overexpression induces a smooth-swimming phenotype (36).

The same abolishment of chemotaxis was observed when Aer-2 (McpB) was overexpressed in both organisms (36). This could be related to its function of being necessary for the complexation of the other Che proteins located in the same operon *che2* (49). Along with this necessity, Aer-2 has a C-terminal extension that specifically allows CheR2, and only CheR2, to methylate it (50), making it appear that the entire Che2 pathway exists to transduce the signal sensed by Aer-2. In the current study we found that deletion of either *cttP* (*mcpA*) or *ae-r2* (*mcpB*) in combination with aer.g1 abolished energy-taxis indicating that there is a connection between the receptors from the *che2* operon and energy-taxis. As these receptors are only present in a narrow subsection of *Pseudomonas* species (the aeruginosa subgroup), which varied widely in their number of Aer homologs, it is unlikely that this is a direct connection. The *in vitro* biochemical studies of Aer-2 indicating that it binds oxygen and the demonstration that its expression in *E. coli* causes a repellent response to oxygen (10), along with the original
characterization of Aer-2 as an aerotaxis receptor (8), strongly indicate that it is a true aerotaxis receptor that binds oxygen. This fits with the observations we made in *P. pseudoalcaligenes* KF707 where the deletion of *aer.g1* and *aer-2* resulted in an energy-tactic negative phenotype. In the soft agar energy-taxis swim plate assay used in this study, oxygen as well as metabolizable carbon source would be diminished at the center of the growth ring. This would allow the energy-sensing of Aer and oxygen sensing of Aer-2 to work together, e.g. in the Aer.g1 and Aer.g2/Aer.g4 mutants that did not display diminished energy-taxis.

**Conclusions**

In *Pseudomonas* Aer does not simply refer to a single receptor for energy-taxis, but to a family of receptors of mostly unknown function. The most ancestral Aer, here named Aer.g1, is indeed an energy-taxis receptor, though other proteins may contribute to its function, including other Aer family members. In *P. pseudoalcaligenes* KF707, unique among *Pseudomonas* species for having 3 Aer homologs, Aer-2 and CttP, also uses these other receptors for energy-taxis. Though as most species do not have these receptors and other evidence indicates their true purpose this may be a minor function. Apart from Aer.g1, Aer family members, Aer-2 and CttP have likely been horizontally transferred within the genus indicating that all these receptors play important roles in *Pseudomonas* physiology.
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-
Acids, 100 bootstrap replicates, JTT amino-acid substitution model, BEST tree topology search operation, tree topology, branch length and rate parameters were optimized. Consense was then used to generate a consensus tree from the bootstrap replicates using the majority rule (extended), and only treating the tree as rooted when the non-pseudomonas root was included.

**Sequence Harmony and Multi-Relief Analysis to Determine Groups**

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

**Comparison of Groups**

Unique regions of each group of Aer homologs were identified by comparing the SHMR scores for each pair-wise comparison with the overall conservation score of each AA and the domain architecture of the Aer protein. Conservation scores were obtained from Jalview (18), and 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$_2$HPO$_4$, 3; NaH$_2$PO$_4$, 1.15; NH$_4$Cl, 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.
Generation of Deletion Constructs and Mutants

Nucleotide sequences for Aer.g1, Aer.g2, Aer.g4, Aer-2 and CttP were obtained from the draft genome sequence of *P. pseudoalcaligenes* KF707 (26). Primers for a ~500bp region up and downstream of each region were generated using Primer BLAST (27). 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 (28). 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(29), 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 (28). White colonies were picked from LB X-Gal gentamycin (20µg/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 (30) 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 (5g/L glucose, minimal salts media) 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.
Chemotaxis Swim Plates

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

Funding Information

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

Acknowledgements

The authors would like to thank 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 and her insights into KF707 physiology and genomic interpretation.
References


49. Güvener ZT, Tifrea DF, Harwood CS. 2006. Two different Pseudomonas aeruginosa


**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.
Supplementary Figure 3: Empirical cumulative distribution function of amino acid multi-relief scores for pair-wise comparisons of adjusted 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. All comparisons show that between 25% and 25% of amino acid positions are above the cutoff.

Supplementary Figure 4: Photographs of energy-taxis swim plates after 24h of growth at 30°C in 50mM pyruvate. 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.

Supplementary Figure 5: 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 6: Mean raw and normalized growth rates of energy-taxis swim-growth-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.

Supplementary Figure 7: 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.

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

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

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.
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.