Evolutionarily stable gene clusters shed light on the common grounds of pathogenicity in the *Acinetobacter calcoaceticus-baumannii* complex

Bardya Djahanschiri[1], Gisela Di Venanzio[2], Jesus S. Distel[2], Jennifer Breisch[3], Marius Alfred Dieckmann[4], Alexander Goesmann[4], Beate Averhoff[3], Stephan Göttig[5], Gottfried Wilharm[6], Mario F. Feldman[2] and Ingo Ebersberger[1,7,8*]

1 Applied Bioinformatics Group, Inst. of Cell Biology and Neuroscience, Goethe University Frankfurt, Frankfurt am Main, Germany

2 Department of Molecular Microbiology, Washington University School of Medicine, St Louis, MO 63110, USA

3 Inst. of Molecular Biosciences, Goethe University Frankfurt, Frankfurt am Main, Germany

4 Bioinformatics and Systems Biology, Justus Liebig University Gießen, Gießen, Germany

5 Institute for Medical Microbiology and Infection Control, University Hospital, Goethe University, Frankfurt, Germany.

6 Robert Koch Institute, Project group P2, Wernigerode, Germany

7 Senckenberg Biodiversity and Climate Research Centre (S-BIKF), Frankfurt am Main, Germany

8 LOEWE Center for Translational Biodiversity Genomics (TBG), Frankfurt am Main, Germany

*Correspondence to: ebersberger@bio.uni-frankfurt.de

Abstract

Nosocomial pathogens of the *Acinetobacter calcoaceticus-baumannii* (ACB) complex are a cautionary example for the world-wide spread of multi- and pan-drug resistant bacteria. Aiding the urgent demand for novel therapeutic targets, comparative genomics studies between pathogens and their apathogenic relatives shed light on the genetic basis of human-pathogen interaction. Yet, existing studies are limited in taxonomic scope, sensing of the phylogenetic signal, and resolution by largely analyzing genes isolated from their functional contexts. Here, we explored more than 3,000 *Acinetobacter* genomes in a phylogenomic framework integrating orthology-based phylogenetic profiling and microsynteny conservation analyses. This allowed to delineate gene clusters in the type strain *A. baumannii ATCC 19606* whose evolutionary conservation indicates a functional integration of the subsumed genes. These evolutionarily stable gene clusters (ESGCs) reveal metabolic pathways, transcriptional regulators residing next to their targets but also tie together sub-clusters with distinct functions to form higher-order functional modules. We shortlisted 150 ESGCs that either co-emerged with, or are found preferentially in, the pathogenic ACB clade. They unveil, at an unprecedented resolution, the genetic makeup that coincides with the manifestation of the pathogenic phenotype in the last common ancestor of the ACB clade. Key innovations are the remodeling of the regulatory-effector cascade connecting LuxR/LuxI quorum sensing via an intermediate messenger to biofilm formation, the extension of micronutrient scavenging systems, and the increase of metabolic flexibility by exploiting carbon sources that are provided by the human host. Specifically, we could show that only members of the ACB clade use kynurenine as a sole carbon and energy source, a substance produced by humans to fine-tune the antimicrobial innate immune response. In summary, this study provides a rich and unbiased set of novel testable

hypotheses on how pathogenic *Acinetobacter* interact with and ultimately infect their human host. They disclose promising routes for future therapeutic strategies.

Keywords

gene cluster, kynurenine, carbohydrate metabolism, phylogenomics, biofilm, quorum sensing, synteny, virulence, innate immune response

Author Summary

The spread of multi- and pan-drug resistant bacterial pathogens is a world-wide threat to human health. Understanding the genetics of host colonization and infection is essential for devising novel ways of treatment. *Acinetobacter baumannii*, a nosocomial pathogen ranked top by the World Health Organization in the list of bacteria for which novel therapeutic approaches are needed, is a prime example. Here, we have carved out the genetic make-up that distinguishes *A. baumannii* and its pathogenic next relatives comprising the *A. baumannii–calcoaceticus* complex from other and mostly apathogenic *Acinetobacter* species. We found a rich spectrum of pathways and regulatory modules that reveals how the pathogens have modified biofilm formation, iron scavenging, and their carbohydrate metabolism to adapt to their human host. Among these, the capability to metabolize kynurenine is particularly intriguing. Humans produce this substance to contain bacterial invaders and to fine-tune the innate immune response. But *A. baumannii* and its allies have turned the table. They feed on kynurenine, and thus most likely also dysregulate human immune response. In summary, our study substantially deepens the understanding of how a highly critical pathogen interacts with its host, which substantially eases the identification of novel targets for innovative therapeutic strategies.

71

72 ## Introduction

73 *Acinetobacter* is a physiologically and biochemically diverse genus of Gram-negative
74 coccobacilli. Most of its species are considered benign, but the genus harbors with the
75 *Acinetobacter calcoaceticus-baumannii* (ACB) complex a group of closely related human
76 opportunistic pathogens [1, 2] that account for the vast majority of severe hospital-acquired
77 *Acinetobacter* spp. infections [3-7]. *Acinetobacter baumannii*, a notorious nosocomial
78 pathogen, is the most renowned member of the ACB complex and signs responsible for up to
79 5% of the total bacterial infections in hospitals worldwide [8]. By now, antibiotic resistance
80 determinants against virtually all available antibiotics drugs are present in *A. baumannii* [9],
81 and multi- or even pan-drug resistant strains are isolated from 44% of all patients with an *A.*
82 *baumannii* induced infection [10]. At the same time, both the frequency and severity of
83 infections have increased. Recent case studies report mortality rates of up to 70% [11-13] as
84 well as growing numbers of (clonal) epidemic outbreaks [14]. Moreover, community-acquired
85 infections by members of the ACB complex begin to rise [6, 15]. In recent years, significant
86 advancements in the molecular characterization of drug resistance mechanisms have led to
87 more informed drug administration schemes for hospitalized patients [6]. Still, the ease with
88 which *A. baumannii* acquires resistance to novel antibiotics [10] makes it likely that resistant
89 strains and their resistance determinants are going to spread at a faster pace than novel
90 antimicrobials become available [16]. As a consequence, *A. baumannii* ranks top in the WHO
91 charts of pathogens for which drug development is most urgent [17].

92 A systemic understanding of how *A. baumannii* interacts with and infects their human host
93 can lead to novel paths for antimicrobial treatments [18-20]. Three main approaches have
94 been used to elucidate the molecular basis of *Acinetobacter* virulence. Candidate approaches

95 (i) have scanned for virulence factors previously characterized in other bacterial pathogens [4,
96 7, 21-25]. To extend the scope beyond pre-compiled virulence factor catalogs, genome-wide
97 approaches (ii) have assessed, for example, the effect of gene knockouts on the infection
98 process (e.g. [12, 26]), or have identified the transcriptional changes under conditions the
99 bacterium typically encounters in the human host (e. g. [23, 27]). These studies are, in theory,
100 the most comprehensive way to identify virulence-related traits. However, experiments are
101 usually performed only on a small set of model strains (e.g. [28]). Moreover, the limited set of
102 tested conditions cannot reflect the diversity of infection sites in the human body.
103 Consequently, results are hard to generalize. Lastly, factors that contribute only indirectly to
104 virulence are hard to pinpoint. For example, we have only begun to understand how
105 adaptations in the carbohydrate metabolism contribute to the virulence of *A. baumannii* [29].
106 Comparative genomics studies (iii) provided complementary evidence with the potential to
107 overcome limitations imposed by the experimental procedures. They aim at identifying
108 evolutionary changes that correlate with the emergence of pathogenic species [30]. To date,
109 the genus *Acinetobacter* encompasses 72 (validly) named and mostly nonpathogenic species
110 [31], which have been isolated from a broad range of habitats such as floral nectar, tree bark,
111 seawater, but also animals [32]. This diversity provides an almost perfect setup to identify
112 changes in the gene repertoires that are associated with the evolutionary emergence of the
113 pathogenic potential, which eventually manifested in the ACB complex. Thus far, comparative
114 genomics studies have begun to shed light on the general evolution of the genus [9] and the
115 clonal epidemiology [33] of *A. baumannii*. They indicated that a major driver of *A. baumannii*'s
116 success as a pathogen is its remarkably flexible genome [9]. High mutation rates [34] paired
117 with the ability to acquire new, or alter the structure or expression of existing genes [11-13]
118 promote a rapid adaptation to novel and adverse environmental conditions. Moreover, its

119 high genomic plasticity [34-36] facilitates the acquisition of antimicrobial resistance
120 determinants complementing the intrinsic *A. baumannii* resistance [37]. However, against the
121 intuition, most of the known virulence determinants were found also in the nonpathogenic
122 members of the genus (e.g. [4, 38-40]). Thus, it is still largely unknown which genetic changes
123 correlate with the emergence of the ACB complex as opportunistic human pathogens, and the
124 genetic basis underlying the adaptation of *A. baumannii* to the human host largely remains to
125 be understood [41].
126 Here, we exploit the availability of thousands of *Acinetobacter* spp. genomes in the public
127 databases (NCBI RefSeq; [42]) to shed light on the evolution of the pathogenic ACB complex
128 at a resolution that extends from a genus-wide overview to the level of individual clonal
129 lineages within *A. baumannii.* For the first time, we integrate genus-wide ortholog searches,
130 with analyses of gene order conservation providing a highly-resolved view on the joint
131 evolutionary fate of neighboring genes using the type strain ATCC 19606 as a reference. This
132 revealed 150 evolutionarily stable gene clusters ($ESGC_{ACB}$) that are prevalent in the ACB
133 complex and rare or absent in the other members of the genus. The functional annotations
134 of these ESGCs provide insights into the genetic and functional specifics of a clade
135 comprising mostly pathogens, and thus direct the focus to key processes likely relevant for
136 the adaptation of the bacterium to the human host. We find that the ACB complex acquired
137 novel genetic modules for the regulation and formation of biofilms, for the scavenging of
138 micronutrients, and have substantially extended their capabilities to exploit a diverse set of
139 carbon sources. Most notably, we show that members of the ACB complex possess the
140 kynurenine (Kyn) pathway and can grow on kynurenine as a sole carbon and energy source.
141 The implications beyond the tapping of a novel carbon source are compelling. Both
142 metabolites are produced by the human host via tryptophan degradation upon bacterial

143 infection and they contribute to the finetuning of the human innate immune response [43,
144 44]. Thus, the presence of the Kyn pathway allows members of the ACB complex to feed on
145 substances produced by the host to defend against bacterial attack. By that, they potentially
146 interfere with the kynurenine-dependent finetuning of the human innate immune response.

147 Results

148 The ability to infect humans emerged in the course of *Acinetobacter* spp. evolution and is a
149 hallmark of the ACB complex. Here, we exploited the full diversity of *Acinetobacter* genomes
150 available in the public databases that was available at the onset of the study (S1 Table). We
151 use this resource to trace changes in the *Acinetobacter* pan-genome that correlate with the
152 manifestation of pathogenicity in the ACB complex. To make the analyses computationally
153 tractable we devised a two-stage strategy.
154 In the priming stage, we determined the evolutionary relationships within the *Acinetobacter*
155 pan-genome with the orthology inference tool OMA [45]. Because the computational
156 complexity of the OMA ortholog search scales exponentially with the numbers of genes in
157 the pan-genome, we compiled a representative set of strains (Set-R) for this analysis. In
158 brief, we considered all available type, reference and representative genomes, as well as all
159 validly named species for which a genome sequence was available at the study onset. This
160 set was filled to a total number of 232 strains by adding further genomes to maximize the
161 phylogenetic diversity of the taxon set (see Methods, and S1 Text: Overview of genome
162 assemblies in Set-R). The corresponding strains together with genome assembly statistics,
163 and, where available, the origins of the isolates are summarized in S2 Table. Members of the
164 ACB complex harbor, on average, 14% more genes than other members of this genus
165 (students t-test - $p<0.001$; S1 Fig A). Gene counts were highly correlated with genome
166 lengths (spearman, $\rho=0.98$), and neither the difference in genome length nor the number of

167 encoded genes was significantly correlated with the assembly status (Completeness status
168 "Complete" vs. Others, Kruskal-Wallis, p=0.311) (see S1 Fig B for further information). The
169 Set-R pan-genome comprises 22,350 orthologous groups harboring 783,306 proteins; 16,000
170 proteins remained singletons (see S1 Fig C for a graphical representation). Rarefaction
171 analyses revealed that the pan-genomes of the entire genus, the ACB complex, and *A.*
172 *baumannii* are open (S1 Text: Pan-genome reconstruction). 889 genes represent the core
173 genome of *Acinetobacter* (S3 Table; S1 Text: Core-genome reconstruction). Eventually, we
174 tentatively annotated gene function in the Set-R pan-genome by linking the individual genes
175 to COGs [46], to KEGG KOs [47], to entries in the virulence factor databases PATRIC [21] and
176 VFDB [48], and by predicting their subcellular localization. To make this data available to the
177 community, we developed the web application Aci-Dash (https://aci-
178 dash.ingress.rancher.computational.bio/; Fig 1), the first web-based platform to interactively
179 browse and explore the *Acinetobacter* pan-genome.
180 In the extension stage, we used a targeted ortholog search to complement the orthologous
181 groups from the SET-R analysis with sequences from the remaining 2,820 *Acinetobacter*
182 genomes (Set-F).

183 ## Consistency-based phylogeny of the genus *Acinetobacter*

184 To establish a stable phylogenetic backbone for our analysis, we reconstructed the
185 maximum likelihood evolutionary relationships of the taxa in Set-R and Set-F, respectively,
186 from three non-overlapping partitions of the 889 core genes. The majority-rule consensus
187 phylogenies from the three trees each (Set-R – Fig 2A; Set-F – Fig 2B; for higher resolution
188 versions see S2 Fig and S3 Fig, respectively) reveal that all named species (at the time of
189 download) as well as the members of the ACB complex are consistently placed into
190 monophyletic clades. Incongruencies between the three partition trees are confined to the

191 branching order within individual species, and here mainly within the densely sampled *A.*
192 *baumannii* and *A. pittii*. This indicates that genetic recombination, which is most likely the
193 source of the incongruent phylogenetic signal [49], is common enough only within species to
194 interfere with phylogenomic reconstructions based on hundreds of genes [50]. Across the
195 genus, we detected and corrected individual taxonomic assignments that are at odds with
196 the phylogenetic placement of the taxa, and most likely indicate mislabeled strains ([51,
197 52]). Specifically, we corrected 16 of such instances within the ACB clade, of which ten were
198 wrongly classified as *A. baumannii* according to NCBI RefSeq. In turn, 60 out of 182 genomes
199 with an unknown taxonomic assignment were placed within the ACB clade (see S4 Table for
200 species and clade assignments including the corrections). Interestingly, a comparison of the
201 average pair-wise nucleotide identity (ANI) across all genomes within Set-R revealed that at
202 least two genomes placed into the ACB clade cannot be associated with any known species
203 (ANI <95%, c. f. S4 Fig). This indicates that the full species diversity of the ACB complex is not
204 yet fully charted.
205 To ease the integration of the phylogenetic information into the following sections, we used
206 one species each to name the individual clades in the *Acinetobacter* phylogeny (Fig 2A).

207 ## Lifestyle and host switches during *Acinetobacter* evolution

208 The two earliest branching clades, named after *A. qingfengensis* (QI) and *A. brisouii* (BR),
209 respectively, solely comprise environmental species. *A. apis*, which was isolated from bees
210 [53], appears as an exception. However, *Acinetobacter* species are sporadically observed in
211 the bee gut, though they are not considered part of the gut microbiome [54]. Instead, they
212 likely represent environmental bacteria that were taken up by the bee with the food [55].
213 Thus, the capability to colonize animals evolved later and most likely in the ancestral species
214 prior to the split of the *A. lwoffii* (LW) clade (Fig 2B). Usually, members of the LW clade are

215 nonpathogenic. Repeated cases of human infection have only been reported for individual
216 strains that mainly group with *A. lwoffii* and *A. radioresistens*, which manifested in vascular
217 catheter-induced bloodstream infections with a low mortality rate [56]. Thus, human
218 infection is likely an exception rather than the rule for this clade. Species descending from
219 the most recent common ancestor *A. baumannii* shares with the *A. baylyi* clade (BA, in
220 particular *A. ursingii*) and with the *A. haemolyticus* clade (HA), respectively, are found
221 increasingly often associated with human pathogenicity, suggesting a progressive adaptation
222 to humans as a host [57]. Members of the monophyletic ACB complex (ACB clade), which
223 subsumes the *A. pittii clade* (PI), the *A. nosocomialis* clade (NO), and the *A. baumannii* clade
224 (B), are all potentially pathogenic, although *A. calcoaceticus* has, thus far, very rarely been
225 seen in the context of human infection. It can be speculated that the few reported cases
226 were due to a miss-classified strain from a different species (see subsection "Consistency-
227 based phylogeny of the genus Acinetobacter" above). Thus, *A. calcoaceticus* at least
228 substantially reduced its pathogenic potential if not lost it completely [25].

229 ## Functional innovation in the Set-R pan-genome

230 We next scanned the Set-R pan-genome for gains of function on the lineage towards
231 contemporary *A. baumannii*. We tentatively stratified the pan-genome by assigning each
232 orthologous group to the inner node in the *Acinetobacter* phylogeny that represents the last
233 common ancestor (LCA) of the two most distantly related taxa in that group (Fig 3A). For
234 each node, we then determined the set of significantly enriched gene functions using the
235 gene ontology (GO) terms from the sub-ontology Biological Process. Because genes with a
236 sparse phylogenetic distribution cannot drive the shared phenotype of a clade, we confined
237 the GO term enrichment analysis to only the subset of orthologous groups, where orthologs
238 were detected in at least half of the subsumed taxa of a node (Fig 3B; Table 1; S5 Table). This

239 revealed several processes with potential relevance for pathogenicity, e.g. cell adhesion,
240 siderophore biosynthesis, and response to oxidative stress. However, reproducing previous
241 observations [4, 38-40], the corresponding genes were assigned to nodes in the phylogeny
242 that predate the emergence of the ACB clade. Only three significantly enriched GO terms
243 were assigned to the node representing the LCA of the ACB clade, and no term was
244 significantly enriched in the genes private to *A. baumannii*.
245

246 Table 1 Selection of overrepresented GO terms in the assigned gene sets of the inner nodes.[1]

Node	GO	Biological Process	p_cor[2]	RT[3]	RB[4]	d[5]	SC[6]	SA[7]	%B[8]	%ACB[9]	%nACB[10]
ACB+BR (n=230)	GO:0018189	pyrroloquinoline quinone biosynthesis	$<10^{-7}$	8.6	5.0	9	192	5	97.8	98.7	17.0
	GO:0006855	drug tm transport	$<10^{-7}$	2.4	9.5	5	53	1	96.4	96.8	23.4
	GO:0015810/3	aspartate/L-glutamate tm transport	$<10^{-7}$	4.3	2.5	10	96	2	96.4	97.8	24.8
	GO:0071705	nitrogen compound transport	$<10^{-7}$	2.8	9.1	4	62	1	100	100	27.7
	GO:0019439	aromatic compound catabolism	$5.2*10^{-3}$	2.2	14.0	4	49	2	92.7	95.7	28.4
	GO:0006631	fatty acid metabolism	$<10^{-7}$	3.3	6.5	7	73	1	89.1	91.4	36.2
	GO:0019619	3,4-dihydroxybenzoate catabolism	$<10^{-7}$	3.2	10.6	8	72	1	90.9	92.5	39.0
	GO:0007155	cell adhesion	$<10^{-7}$	4.6	10.6	2	103	1	98.2	98.9	40.4
	GO:0010124	phenylacetate catabolism	$<10^{-7}$	10.4	8.5	8	233	4	89.6	90.6	45.7
ACB+LW (n=223)	GO:0031388	organic acid phosphorylation	$<10^{-7}$	2.9	2.4	6	54	1	98.2	98.9	14.9
	GO:0009437	carnitine metabolism	$3.1*10^{-3}$	0.2	0.1	6	3	1	94.5	88.2	22.0
	GO:0071705	nitrogen compound transport	$<10^{-7}$	4.4	10.5	4	83	2	99.1	94.6	23.4
	GO:0036104	Kdo2-lipid A biosynthesis	$<10^{-7}$	2.4	3.8	8	45	1	94.5	91.4	29.8
	GO:0009435	NAD biosynthesis	$<10^{-7}$	3.4	15.7	11	65	1	89.1	91.4	30.5
	GO:0009116	nucleoside metabolism	$<10^{-7}$	3.5	13.0	6	67	1	96.4	95.7	31.2
	GO:0055085	tm transport	$<10^{-7}$	24.2	149.4	4	459	7	90.2	89.1	38.7
	GO:0019557/6	histidine catabolism to glutamate and formate/formamide	$<10^{-7}$	9.9	6.1	10	188	3	95.8	97.5	41.6
	GO:0006351	transcription, DNA-templated	$<10^{-7}$	86.9	320.1	9	1645	21	95.5	93.7	41.8
	GO:0019290	siderophore biosynthesis	$<10^{-7}$	8.0	4.8	8	170	3	95.8	72.8	41.8
ACB+BA (n=152)	GO:0045150	acetoin catabolism	$<10^{-7}$	33.6	0.8	7	36	1	100	100	14.9
	GO:0019290	siderophore biosynthesis	$<10^{-7}$	29.9	5.6	8	32	1	96.4	97.8	22.0
	GO:0006351	transcription, DNA-templated	$<10^{-7}$	76.5	0.03	9	82	2	84.5	84.9	26.2
ACB+HA (n=143)	GO:0022904	respiratory electron transport chain	$<10^{-7}$	45.9	5.39	5	44	0	90.9	79.6	11.3
	GO:0006351	transcription, DNA-templated	$<10^{-7}$	78.3	314.8	9	75	0	90.9	91.4	27.0
ACB (n=93)	GO:0009372	quorum sensing	$<10^{-7}$	8.8	0.3	3	13	2	86.4	90.3	0.0
	GO:0006351	transcription, DNA-templated	$<10^{-7}$	98.9	318.1	9	146	5	92.0	87.1	0.0
	GO:0006979	response to oxidative stress	$4.7*10^{-5}$	9.5	17.4	3	14	1	89.1	78.5	0.0

247

[1] The full list of enriched terms is provided in S5 Table 1.
[2] corrected p-value
[3] Ratio of genes with term in test set of that node ($\times 10^{-3}$)
[4] Ratio of genes in background set of that node ($\times 10^{-4}$)
[5] Depth of term in GO tree
[6] Count of subjects with term in study set of that node
[7] Number of ortholog groups associated with proteins in the study set. For this table, redundancy was reduced by selecting the lowest depth GO terms if any two GO terms had non-empty intersections between their sets of protein sequence identifiers or associated ortholog group identifiers.
[8] Mean prevalence of associated ortholog groups in *A. baumannii*
[9] Mean prevalence of associated ortholog groups in the ACB clade
[10] Mean prevalence of associated ortholog groups in *Acinetobacter* spp.

248 ## Evolutionary and functional units in the ATCC 19606 gene set

249 The GO term enrichment analysis revealed only a weak signal for the gain of biological
250 processes that can be directly connected to bacterial virulence in the evolutionarily younger
251 nodes of the *Acinetobacter* phylogeny. However, two factors confound the analysis. Only
252 about two-thirds of all unique sequences in the *Acinetobacter* spp. pan-genome is annotated
253 with a GO term and many of the genes with only unspecific terms. Moreover, the sporadic
254 presence of orthologs to individual genes in taxa outside the ACB clade, e.g. as a result of
255 recombination, confounds the phylogenetic stratification of the pan-genome. To increase
256 both power and resolution, we widened the focus and traced the emergence and evolution
257 of gene clusters in *Acinetobacter* spp. using the type strain *A. baumannii* ATCC 19606 as a
258 reference. For each gene in the type strain, we integrated its node assignment, the
259 abundance of orthologs within *A. baumannii*, within other members of the ACB clade, and
260 within the remaining taxa in SET-R, respectively, and, where applicable, the overrepresented
261 GO terms (Fig 4). As a recurrent theme, we observed that genes linked to the same
262 overrepresented GO term reside adjacent to genes with highly similar phylogenetic
263 abundance patterns, despite being occasionally assigned to different nodes in the
264 phylogeny. We subsequently connected neighboring genes with significantly similar
265 abundance patterns across the entire genome of *A. baumannii* ATCC 19606 to form
266 candidate clusters. A candidate cluster was then propagated to represent an evolutionarily
267 stable gene cluster (ESGC), which likely forms a functional entity if its gene order was found
268 conserved within the ACB clade (S6 Table 3). Subsequently, we shortlisted 150 $ESGC_{ACB}$ that
269 are abundant among members of the ACB clade and, at the same time, rare or even absent
270 outside the clade (SE1.3 and SE1.4; see S1 Text: Workflow of ESGC identification and
271 shortlisting). As the last step, we manually validated the automatic $ESGC_{ACB}$ assignment for

272 44 clusters whose sets of genes were each functionally annotated to a level that allowed us
273 to infer general cluster function (Fig 5), and for further 10 clusters with an unknown function
274 but that are almost exclusively found in the ACB clade (Fig 6). These 54 $ESGC_{ACB}$ can be
275 broadly distinguished into four categories: The cluster comprises (i) a known operon or
276 metabolic gene cluster (e.g., $ESGC_{ACB}$-0622, phenylacetate metabolism), (ii) a group of
277 functionally related genes that are linked to genes, whose functional link, e.g. acting as a
278 transcriptional regulator, was unknown thus far (e.g., $ESGC_{ACB}$-0162), (iii) super-clusters
279 connecting two or more clusters with distinct functions to form a higher-order functional
280 unit ($ESGC_{ACB}$-0394, *adeFGH* and *hisQMP*), and (iv) clusters of genes with an unknown
281 function (e.g., $ESGC_{ACB}$-0503 to 0510; Fig 6). We hypothesize that the adaptation of
282 pathogenic *Acinetobacter* species to their human host is likely mirrored in the various
283 functions that these $ESGC_{ACB}$ convey. In the following sections, we highlight a selection of
284 clusters related to persistence, micronutrient acquisition, and the evolution towards
285 nutritional flexibility (see S1 Text: Gene clusters (ESGCs) not discussed in main manuscript).

286 The evolutionary emergence of the ACB clade coincides with changes in quorum sensing and
287 biofilm formation

288 Quorum sensing and biofilm formation are key determinants of *A. baumannii* virulence [58-
289 60]. Both functions are represented by $ESGC_{ACB}$-0162 (14 genes) and 0410 (8 genes).
290 $ESGC_{ACB}$-0162 represents the regulatory module of this process. It harbors the Lux-type
291 quorum sensing system (QS_{Lux}), which regulates motility and biofilm formation in *A.*
292 *baumannii,* and additionally, a biosynthetic gene cluster containing a non-ribosomal peptide
293 synthetase here referred to as NRPS cluster. Both genes of QS_{Lux}, *abaI* and *abaR,* are
294 separated by a short gene that was tentatively named *abaM* (Fig 7A; see also S1 Text:
295 Characterization of AbaM). This three-gene architecture is conserved in the ACB clade, and it

296 is common in *Burkholderia* spp. [61], where the intervening short gene acts as a negative
297 regulator of the QS_{Lux} system [62]. Recently, initial evidence emerged that AbaM in the *A.*
298 *baumannii* strain AB5075 indeed regulates quorum sensing and biofilm formation [63]. In
299 the light of our results, we propose that AbaM is an understudied modulator of quorum
300 sensing in the entire ACB clade.
301 The NRPS cluster (Fig 7A) produces a three-amino acid lipopeptide, Ac-505 [60], which likely
302 plays a central role in regulating bacterial motility and biofilm formation [64]. Disrupting its
303 biogenesis alters the expression of numerous factors involved in biofilm formation and
304 surface adherence [65], in particular the chaperon-usher pili (CUP) and the archaic
305 chaperon-usher pili (CSU). Consequently, host cell adhesion and virulence of *A. baumannii*
306 are substantially reduced. Here, we provide first-time evidence that the evolutionary fate of
307 the NRPS cluster is intimately intertwined with that of the QS_{Lux} cluster. We found that the
308 rare strain-specific loss of the QS_{Lux}-cluster determines the loss of the NRPS cluster, which
309 implies that they not only form an evolutionary but also a functional unit. Interestingly,
310 strains lacking $ESGC_{ACB}$-0162 are not randomly distributed. Most prominently, the cluster is
311 missing in almost all (48/55) *A. baumannii* strains representing the international clone type
312 (IC) 8 (Fig 5). The formation of a higher-order module comprising the QS genes and an NRPS
313 biosynthetic gene cluster is a repeated scheme during bacterial evolution. For example, the
314 methane-oxidizing bacterium *Methylobacter tundripaludum* harbors an NRPS biosynthetic
315 gene cluster that was integrated between the *abaI* and *abaR* orthologs. And the production
316 of the corresponding extracellular factor is under control of the QS cluster [66]. NRPS-
317 dependent molecules have been implicated to mediate interspecific communications across
318 kingdoms both in symbiotic and pathogenic communities. In *P. aeruginosa*, the interplay of
319 N-acyl-L-homoserine lactone-dependent quorum-sensing signaling and an NRPS-dependent

320 biosynthesis of bacterial cyclodipeptides (CDPs), which act as auxin signal mimics, modulates
321 the communication to its host plant *Arabidopsis thaliana* [67]. It can be speculated that
322 $ESGC_{ACB}$-0162 may similarly coordinate, the communication between the bacteria and their
323 human host.
324 $ESGC_{ACB}$-0410, harbors the Csu cluster responsible for biofilm formation on abiotic surfaces
325 via archaic chaperon-usher pili [68, 69] (Fig 7B) together with a transcriptional regulator of
326 the TetR/AcrR family (TFTRs) (S5 Fig id:0410). TFTRs represent one-component systems that
327 regulate a broad variety of cellular processes in bacteria, among them many that are related
328 to virulence such as efflux pump expression and biofilm formation [70, 71]. Notably, they are
329 often encoded alongside their target operons. To the best of our knowledge, regulation of
330 the Csu cluster via an adjacent TFTR has never been reported. Thus, next to the two-
331 component systems BfmRS [72] and GacSA [73], a third hitherto undescribed one-
332 component system, seems to be involved in regulating the formation of Csu pili.
333 In line with our finding that $ESGC_{ACB}$-0162 and $ESGC_{ACB}$-0410 have similar abundance
334 patterns (cf. Fig 5), the regulation of the Csu cluster appears under the direct control of Ac-
335 505 [26, 65]. Thus, Ac-505 likely acts as a modulator between biofilm formation on abiotic
336 and biotic surfaces. However, contrary to the QS_{Lux} - NRPS supercluster, the Csu cluster was
337 lost multiple times independently in the ACB clade (Fig 5). Given its terminal position in the
338 regulator-effector cascade, this indicates a lineage-specific fine-tuning of biofilm formation.
339 Interestingly, within *A. baumannii*, we find that all 55 IC8 strains in our dataset lack both the
340 QS_{Lux} - NRPS supercluster and the Csu cluster, which points towards substantial changes in
341 the way how IC8 strains regulate biofilm formation.

342 ## KatA – an ACB clade specific catalase

343 The genome of *A. baumannii* ATCC 19606 harbors five putative catalases: *katA*, *katE*, *katE-*
344 *like*, *katG*, and *katX* [23, 74]. Note, that both Sun et al. [74] and Juttukonda et al. [23, 74]
345 refer to a catalase labeled *katE*, which, however, represent different genes (locus tags
346 A1S_1386/A1S_3382 and DJ41_RS22765/DJ41_RS10660 in *A. baumannii* ATCC 17978 and
347 ATCC 19606, respectively). We, therefore, named the latter *katE*-like. *KatA* is the only
348 catalase that is exclusively found in the ACB clade. The corresponding gene resides in a
349 cluster next to a putative MFS transporter and a cytochrome b561 ($ESGC_{ACB}$-0624, Fig 7C).
350 The KatA cluster is highly conserved in all species of the ACB clade except *A. calcoaeceticus*,
351 where it has been lost (cf. Figs 5 and 7C). Upon host infection, both neutrophils and
352 macrophages recruit radical oxygen species (ROS) for bacterial clearance [75, 76], and thus
353 ROS defense mechanisms are an essential contributor to bacterial virulence. However, an
354 initial investigation in *Ab* ATCC 17978 found no obvious link between KatA and ROS
355 protection [23]. Still, the abundance pattern of $ESGC_{ACB}$-0624 is perfectly in line with the
356 hypothesis that this cluster contributes to virulence in pathogenic members of the ACB
357 clade, and more comprehensive studies are needed to elucidate its precise role during
358 bacterial infection.

359 ## Metabolic adaptation – Micronutrient acquisition is refined in the ACB-clade

360 All members of the genus *Acinetobacter* must supply their demand of essential
361 micronutrients, most importantly iron and zinc. Acquisition systems for these two metals are
362 therefore represented in the genus-wide core genome (Fig 8). However, both iron and zinc
363 are actively sequestered by the host to starve invading pathogens, a strategy referred to as
364 nutritional immunity [77]. This likely results in a strong selective pressure for the pathogenic

365 ACB clade to optimize scavenging systems such that the reduced bioavailability of these
366 metal ions in the host can be counterbalanced.
367 Acquisition systems for iron, whose limited availability at the host-pathogen interface is
368 considered one of the key obstacles for invading and persisting within the human host, are a
369 showcase example. The iron transporter system *feoABC* represents the evolutionary core of
370 iron uptake. It is complemented, in many but not all taxa [28] both inside and outside of the
371 ACB clade, by the baumannoferrin cluster (Fig 8). Two further clusters extend the spectrum
372 of iron uptake systems exclusively in the ACB clade. $ESGC_{ACB}$-0498 represents the 2,3-
373 dihydroxybenzoic acid synthesis cluster (*entAB*), which synthesizes a siderophore precursor
374 [78]. $ESGC_{ACB}$-0368 and 0369 together resemble the acinetobactin biosynthesis clusters
375 *bauA-F*, *basA-I* and *barAB*. However, a third cluster, $ESGC_{ACB}$-0485, that very likely represents
376 an ABC-type Fe^{3+}-hydroxamate transport system seems to extend the diversity of iron
377 uptake systems in the ACB clade even further. It encodes a substrate-binding protein, an iron
378 complex ABC transporter (permease), an ATP-binding protein, and an N-Acetyltransferase
379 protein (GNAT family). The AraC-family-like transcriptional regulator, which is located
380 downstream on the opposite strand in $ESGC_{ACB}$-0485, likely controls the expression of this
381 cluster. In line with this operon-like organization, these genes are jointly downregulated
382 under mucin-rich conditions [79]. The complex and seemingly redundant infrastructure for
383 iron uptake in the ACB clade seems at odds with a recent study in *A. baumannii* ATCC 17978,
384 which stated that acinetobactin is the only system that is necessary for *A. baumannii* to grow
385 on host iron sources [80]. However, here we show that this conclusion does not generalize
386 to the entire ACB clade. The pathogens *A. nosocomialis* and *A. seifertii*, for example, lost the
387 acinetobactin cluster (cf. Fig 5). Instead, it is conceivable that the diversity of iron acquisition

388 systems is an adaptation to diverse niches each requiring different strategies of iron
389 scavenging.
390 Given the essentiality of zinc (Zn), it is not surprising to see that also Zn uptake was refined
391 on the lineage towards the ACB clade. The zinc uptake system Znu, including the distal *znuD*
392 gene, which facilitates resistance to human calprotectin-mediated zinc sequestration [23], is
393 evolutionarily old and part of the genus-wide core genome (Fig 8) The histidine utilization
394 (Hut) system (*hutCDUHTIG*, $ESGC_{ACB}$-0215) is prevalent in the ACB clade, though not
395 exclusively. This system ensures the bio-availability of Zn via the histidine catabolism both
396 under high availability and starvation of Zn. However, it requires histidine to be abundant.
397 Interestingly, the most recent acquisition in the Zn metabolism is the putative
398 metallochaperone, ZigA. The corresponding gene resides directly adjacent to $ESGC_{ACB}$-0215,
399 and thus is likely an evolutionary more recent extension of this cluster. *zigA* was found active
400 only under Zn starvation, where it increases the bioavailability of Zn also under histidine
401 depletion [81] and counteracts nutritional starvation.
402 Manganese (Mn^{2+}) is required only in small amounts and is mostly used for coping with
403 reactive oxygen species (ROS), as Mn^{2+}, other than Fe^{2+}, does not promote the Fenton
404 reaction that converts H_2O_2 to highly damaging hydroxyl radicals [82, 83]. Therefore, Mn
405 uptake systems should be prevalent in bacteria frequently exposed to ROS stress,
406 particularly in the pathogenic *Acinetobacter* strains. Thus far, only one Mn acquisition
407 system, *mumRTLUHC* [84], has been identified in *Acinetobacter spp*. This system is
408 represented by $ESGC_{ACB}$-0611 and plays an essential role in protecting *A. baumannii* against
409 calprotectin-mediated Mn depletion by the host and contributes to bacterial fitness in a
410 murine pneumonia model [84]. $ESGC_{ACB}$-0611 is found throughout the genus though less
411 frequently in clades that are more distantly related to the ACB clade (cf. Fig 5 and 8). Within

412 the HA-clade, several species including *A. tjernbergiae*, *A. junii*, *A. beijnerickii,* and *A.*
413 *haemolyticus* lack both the putative manganese transporter gene *mumT* and the gene
414 encoding a putative hydrolase *mumU* (S5 Fig id:0611). Within IC 3, several strains lack the
415 entire cluster. These taxa either found alternatives to Mn^{2+}-dependent processes for coping
416 with oxidative stress, are more vulnerable to ROS, or they scavenge Mn^{2+} via a mechanism
417 that is still hidden in functionally uncharacterized gene clusters.
418 In summary, we see a clear signal that the ACB clade is enriched for genes and gene clusters
419 that functionally complement the genus-wide available and evolutionarily old metal uptake
420 systems. In line with the reinforcement hypothesis, these more recently acquired clusters
421 seem particularly important for metal scavenging during infection, i.e. when the metals are
422 actively sequestered by the host [80, 81].

423 ### Carbohydrate metabolism - Evolution towards nutritional flexibility

424 The ability of individual *Acinetobacter* strains to utilize a broad spectrum of carbon sources is
425 important for their adaptation to different environments, including the human host [40, 85-
426 88]. However, it is largely unknown when the corresponding metabolic pathways were
427 acquired during *Acinetobacter* evolution, how widespread they are, and if and to what
428 extent they are connected to the pathogenicity of the ACB clade. More than 20 of the
429 shortlisted $ESGCs_{ACB}$ represent pathways that shuttle metabolites into the carbohydrate
430 metabolism of the bacterium (Fig 5, in red font, and see Fig 9 for a selection), many of which
431 are prevalent in the human body. The corresponding gene clusters mostly channel these
432 metabolites into catabolic processes (see below). However, the genes involved in the
433 glucose/gluconate metabolism seem to fuel anabolic processes.

434 ## Glucose / Gluconate metabolism

435 Glucose and gluconate serve as carbon and energy sources for individual species in the
436 genus *Acinetobacter*, e.g., *A. soli, A. apis,* and *A. baylyi.* For *A. baylyi* ADP1 it was shown that
437 the glucose catabolism involves the Entner-Doudoroff pathway [89]. Members of the ACB
438 clade have lost the ability to use glucose as a carbon source [90] (see also S6 Fig C and S7
439 Table). It is, thus, surprising that we find the genetic infrastructure to feed both molecules
440 into the bacterial metabolism almost exclusively in the ACB clade.

441 $ESGC_{ACB}$-0112 comprises the gluconate permease (GntP) that shuttles gluconate from the
442 periplasm into the bacterial cell (Fig 9, yellow pathway). The cluster further encodes the
443 kinase GntK, which phosphorylates gluconate into 6-phosphogluconate, and the enzymes
444 Edd and Eda of the Entner-Doudoroff pathway, which link to the pentose phosphate
445 pathway that produces pyruvate. Members of the ACB clade also possess two variants of a
446 glucose dehydrogenase (*gdh*), which catalyze the reaction from D-glucose to D-gluconate in
447 the periplasm [91]. The membrane-bound variant (*gdhA*), forms together with an outer
448 membrane porin a cluster of two genes, $ESGC_{ACB}$-0287 (cf. S1 Data id:0287), which is
449 ubiquitous across *Acinetobacter* spp. We note that the porin is orthologous to OprB in *P.*
450 *aeruginosa*, where it facilitates the diffusion of various sugars - including glucose - into the
451 periplasm. The second, soluble Gdh (*gdhB*) is confined to and nearly ubiquitous in the ACB
452 clade (S7 Fig, S6 Table 1, id:HOG3408).

453 The prosthetic group for both Gdh, pyrroloquinoline quinone (PQQ), is a small, redox-active
454 molecule that serves as a cofactor for several bacterial dehydrogenases. $ESGC_{ACB}$-0497
455 comprises six genes that together represent the PQQ biosynthesis pathway: *pqqABCDE* and
456 an additional membrane-bound dipeptidase referred to as *pqqF* in *Klebsiella pneumoniae*
457 [92] (cf. Fig 9, pink). All genes reside contiguously on the same strand suggesting an operonic

458 structure. The complete cluster is present in almost all genomes of the ACB clade, although
459 we had to manually confirm the presence of *pqqA*, because its length (40 amino acids) is
460 below the length cutoff of the ortholog assignment tool. Only 31 out of 2436 strains in the
461 ACB clade have lost the ability to synthesize PQQ, among them the model strain *A.*
462 *baumannii* ATCC 17978. Cluster abundance outside the ACB clade is low (<20%), but it is
463 present in all strains of species with demonstrated ability to assimilate glucose and
464 gluconate (i.e. *A. soli*, *A. baylyi*, and *A. apis*, cf. [90]; SE2.0497).
465 The holoenzymes *GdhA* and/or *GdhB*, in theory, could establish a gapless route for glucose
466 via this gluconate 'shunt' into the cell for further degradation via the Entner-Doudoroff
467 pathway, even in the absence of a dedicated Glucose transporter. Why then does none of
468 the tested strains in the ACB clade grow with glucose as sole carbon and energy source? We
469 hypothesize that they utilize this route for anabolic processes exclusively, e.g. for the
470 production of polysaccharides as it was demonstrated for *P. aeruginosa* [93].

471 Carbohydrate Catabolism

472 The ACB clade has substantially increased its repertoire of catabolic pathways for alternative
473 carbon sources compared to taxa outside this clade [94]. For a small number of mostly hand-
474 picked *A. baumannii* strains, previous studies have experimentally confirmed the ability to
475 grow on tricarballylate and putrescine, malonate, butanediol and acetoin, phenylacetate,
476 muconate, glucarate, galactarate (mucate), and 4-hydroxyproline as sole carbon sources [40,
477 85]. Our analyses identified the corresponding gene clusters among the $ESGC_{ACB}$. Hence, the
478 ability to use these resources is prevalent in the ACB clade, whereas non-ACB species have
479 to rely largely on different carbon sources. We will highlight two examples that likely
480 represent an adaptation to humans as a host.

481 D-glucarate (saccharate) is a major organic acid in human serum [95]. $ESGC_{ACB}$-0568
482 comprises all necessary genes for glucarate and galactarate (mucic acid) degradation (Fig 9).
483 In *Salmonella enterica* serovar Typhimurium deletion of the D-glucarate/D-galactarate
484 permease ortholog attenuated virulence [96]. Further, galactarate digestion was shown to
485 increase the colonization fitness of intestinal pathogens in antibiotic-treated mice and to
486 promote bacterial survival during stress [97]. Given that $ESGC_{ACB}$-0568 is almost exclusively
487 confined to the ACB clade, it is conceivable that this cluster contributes to colonization and
488 virulence in pathogenic *Acinetobacter* species. It is therefore interesting that within *A.*
489 *baumannii* the cluster is almost absent in IC2 strains (0.08% prevalence in Set-F; *cf.* Fig 5).
490 Carnitine is essential for the oxidative catabolism of fatty acids in humans [98]. $ESGC_{ACB}$-0016
491 comprises six genes necessary for catabolizing carnitine [99]. A LysR-type transcriptional
492 regulator likely controls the activity of this cluster. The remaining five genes represent a
493 putative tartrate dehydrogenase (*ttuC*), a BCCT-family carnitine transporter (Aci01347), a
494 generically annotated alpha/beta hydrolase which possibly catalyses the conversion of D-
495 acylcarnitine into L-carnitine (see Fig 9, green), and the genes encoding the two subunits of
496 the carnitine monooxygenase CntA and CntB. The latter two genes are separated by a gene
497 that is tentatively annotated as an NAD-dependent succinate-semialdehyde dehydrogenase.
498 However, two lines of evidence indicate that the precise function of this gene as well as that
499 of the putative tartrate hydrogenase might both differ. In the literature, the putative
500 succinate-semialdehyde dehydrogenase is speculated to act as malic semialdehyde
501 dehydrogenase [99], an enzyme that converts malate semialdehyde into malate. Further,
502 the putative tartrate dehydrogenase belongs to the KEGG orthologous group KO7246 which
503 is annotated as a D-malate dehydrogenase. The product of the latter enzyme, pyruvate, can
504 be further processed into oxaloacetate, which serves as a substrate for the tricarboxylic acid

505 (TCA) cycle (cf. Fig 9). Assuming that the putative malic semialdehyde dehydrogenase
506 produces D-malate, then, in the presence of an appropriate transporter, this cluster should
507 allow the members of the ACB clade to utilize D-malate as a carbon source. We tested this
508 hypothesis, and confirmed that *Ab* ATCC 19606 indeed grows on D-malate (S6 Fig B), which
509 corroborates initial growth experiments [100]. We note that the unusual production of the
510 D-malate enantiomer rather than L-malate would have a further interesting implication. It
511 potentially allows the bacterium to accumulate D-malate in conditions when carnitine is
512 abundant, without interfering with the stoichiometry of the remaining substrates of the TCA
513 cycle.
514 Thus far, two *A. baumannii* strains have been shown to use carnitine as sole carbon source
515 [101]. We evaluated exemplarily that the absence of the cluster indeed correlates with
516 *Acinetobacter* inability to grow on carnitine and is not functionally complemented by an
517 alternative degradation pathway. Both, *A. baylyi* ADP1 and the *A. calcoaceticus* type strain
518 (DSM 30006), which both lack the $ESGC_{ACB}$-0016, did not grow on carnitine after 24h.
519 The clusters abundance profile reveals that the ability to metabolize carnitine occurs also
520 outside the ACB clade (31 strains in Set-R; cf. Fig 5). However, 25 of these strains were
521 isolated from infected patients, 2 from hospital sewage water, and only 3 strains were
522 sampled from the environment (cf. S8 Table). The isolation origin of the one strain is
523 unknown. The presence of the carnitine cluster, therefore, correlates surprisingly well with
524 the pathogenic potential of a strain and it will be interesting to test a causal dependence. In
525 support of causality, we find that the carnitine cluster is absent in *A. calcoaceticus*, the only
526 species of the ACB clade that is nonpathogenic or at least has substantially reduced virulence
527 (cf. S1 Text: $ESGC_{ACB}$-0016).

528 Novel carbon sources

529 The functional annotations of the $ESGC_{ACB}$ in Fig 5 indicate that the list of potential carbon
530 sources for *A. baumannii* and other members of the ACB clade is still incomplete. We find
531 degradation pathways for xanthine, 2-aminoethylphosphonate, acyclic terpenes,
532 vanillin/valinate, taurine, and anthranilate (see S1 Text: Gene clusters (ESGCs) not discussed
533 in main manuscript). To the best of our knowledge, none of these metabolites have been
534 considered as potential carbon sources for *A. baumannii,* although a putative Xanthine
535 dehydrogenase has been characterized previously via heterologous expression in *E. coli*
536 [102]. One cluster is, however, particularly interesting because it intertwines the bacterial
537 carbohydrate metabolism with the human tryptophan catabolism. $ESGC_{ACB}$-0452 (Fig 10A)
538 harbors the kynurenine hydrolase KynU, which catalyzes the cleavage of kynurenine to
539 anthranilic acid and alanine (Fig 10B). Within the same cluster, we identified an AsnC-type
540 transcriptional regulator, a putative amino acid permease, and a gene generically annotated
541 as an alpha/beta-hydrolase. Notably, this hydrolase is listed as an ortholog to the human
542 kynurenine formamidase (KynB) in the OMA database [103] with which it shares the same
543 domain architecture (Fig 10C). Together with its evolutionarily stable localization in the
544 vicinity of KynU, this provides convincing evidence that the *A. baumannii* hydrolase
545 resembles a hitherto overlooked KynB. This enzyme transforms formyl-kynurenine into
546 kynurenine and, thus, acts immediately upstream of KynU in the Kynurenine pathway of
547 tryptophan degradation (*cf.* Fig 10B). A similar gene cluster was recently described in *P.*
548 *aeruginosa* (Pae) [104, 105], however with two notable exceptions: $KynB_{Aba}$ is substantially
549 more similar both in sequence and domain architecture to the human KynB than the $KynB_{Pae}$
550 (Fig 10C), and we found no trace of an enzyme that catalyzes the formation of N-formyl-
551 kynurenine from tryptophan in *A. baumannii*.

552 We next confirmed that the presence of the kynurenine cluster allows *A. baumannii* to grow
553 on kynurenine as a sole carbon- and energy source (Fig 10D). A deletion of the Kyn-cluster
554 abolished growth on kynurenine but not on anthranilate, the product of KynU (Fig 10D). In a
555 last step, we investigated the phylogenetic profile of the Kyn-cluster in greater detail (Fig 10E
556 and S2 Data). Within *Acinetobacter*, the cluster is almost exclusively present in the
557 pathogens of the ACB clade, and it has been lost several times independently within *A.*
558 *calcoaceticus*. Along the same lines, we find the Kyn-cluster in a proteobacteria-wide screen
559 across more than 1,000 species (see S9 Table for the list of genomes and identifiers) only in a
560 few taxa, of which many are opportunistic human pathogens (SF5). Taken together, our
561 findings provide strong evidence that members of the ACB clade possess the genetic
562 infrastructure to interfere with the tryptophan metabolism of humans, as it was already
563 shown for *P. aeruginosa* [105]. This opens up a novel and hitherto unexplored route of how
564 these pathogens interacts with its human host.

565 ## Discussion

566 In contrast to more virulent bacteria, the opportunistic nosocomial pathogens from the
567 genus *Acinetobacter* pursue a resist-and-persist strategy [25]. Instead of pinpointing
568 individual key determinants of bacterial virulence, this requires the unraveling of a likely
569 broad and less specific genetic basis conveying adaptation to clinical environments and, at
570 the same time, to the human host. Experimental studies, either *in vitro* bringing the
571 advantage of controlled experimental conditions, or *in vivo* with the advantage of a realistic
572 infection model have provided fundamental insights into the pathobiology of *Acinetobacter*
573 [19, 20, 40]. However, pathogens encounter – and therefore have to adapt – to diverse
574 environments during host infection [106]. The resulting selective landscape is complex, and
575 therefore hard to reproduce in an experimental setup. Virulence determinants that are

576 relevant only under specific conditions or whose functions blend in with the bacterial
577 metabolism are easy to miss. Evolutionary approaches can close this gap. They focus on the
578 signal if a gene or a gene cluster likely contributes to pathogenicity, reflected by their
579 prevalence preferentially in pathogens, independent of the precise conditions when it is
580 active. Here, we have charted the genetic specifics of the pathogenic ACB clade at a
581 resolution ranging from a genus-wide overview down to individual clonal lineages within *A.*
582 *baumannii*. To ease the future integration of evolutionary evidence with functional studies,
583 we have developed the *Acinetobacter* dashboard Aci-Dash. This is the first web application
584 that allows the community to mine the abundance profiles of genes encoded in 232
585 representative genomes together with their functional annotations and their connection to
586 virulence factors.
587 Comparative genomics studies across the genus *Acinetobacter* have been performed before
588 (e.g. [9, 107]). The integration of orthology assignments with shared synteny analyses at a
589 scale that spans several thousand *Acinetobacter* genomes is yet unprecedented. The $ESGC_{ACB}$
590 detected here form relevant starting points for further unravelling the regulatory and
591 functional network in a human pathogen. They result in a rich set of testable hypotheses
592 whose experimental validations will likely deepen the understanding of the genetic basis of
593 *Acinetobacter* pathogenicity. Moreover, the lineage specific absence of ESGC within the
594 pathogenic clade, most prominently demonstrated by the loss of the QS-NRPS cluster
595 together with the Csu cluster in the IC 8, or of several clusters in *A. calcoaceticus*, helps to
596 better predict in which characteristics individual strains, clonal lineages, and species differ
597 from the prototype of an *Acinetobacter* pathogen.
598 The functions conveyed by these clusters are diverse and many can be readily connected to
599 *Acinetobacter* pathogenicity: Quorum sensing and biofilm formation [93], ROS response

600 [108], and micronutrient acquisition [77]. On top of these, the abundance of gene clusters
601 involved the carbohydrate metabolism, a largely uncharted area of *Acinetobacter* virulence
602 factors [29], strongly suggests that members of the ACB clade follow a general evolutionary
603 trend towards greater metabolic flexibility, which is common to many bacterial pathogens
604 [30, 106]. Such generalists have a selective advantage over niche-specialists in environments
605 that are frequently disturbed or altered [109].
606 Metabolic interaction with the host, in particular those involving amino acids, have an
607 interesting further implication, that has thus far not been considered in the context of
608 *Acinetobacter* pathogens. They can modulate regulatory systems involved, for example, in
609 the finetuning of the host immune response [30, 110]. The Kynurenine (Kyn) pathway
610 detected in this study is a likely example of this connection. This cluster is almost entirely
611 confined to the pathogenic members of the ACB-clade. Interestingly, and in contrast to the
612 human pathogen *P. aeruginosa,* which also harbors a Kyn pathway, the key enzyme, IDO
613 (KynA), that allows the bacteria to directly metabolize tryptophan via the kynurenine
614 pathway is missing (cf. Fig 10B). How then do they fuel this pathway? Interestingly,
615 tryptophan depletion via the kynurenine pathway is an important human immune defense
616 mechanism upon bacterial infection [111]. Here, we provide first time evidence that
617 pathogenic *Acinetobacter* species from the ACB clade can use the intermediate metabolites
618 of the host response as additional carbon- and energy sources, likely further promoting its
619 growth. However, the ability to degrade the intermediates of the human tryptophan
620 catabolism has a further interesting implication for in the host-pathogen interaction. While it
621 was shown that *P. aeruginosa* produces elevated levels of kynurenine to inhibit ROS
622 production and aid bacterial survival [112], *Acinetobacter* species in possession of the Kyn
623 cluster must pursue a different strategy. By scavenging kynurenine from their environment,

624 they can potentially interfere with the homeostasis of the human immune system, in
625 particular with its suppressive effect on T cells and Natural killer cells. This potentially leads
626 to host tissue damage due to the now uncontrolled production of ROS [76, 79]. Though at a
627 high cost, this would allow the bacteria to use the rich nutrient resources enclosed in the
628 host cells.
629 In summary, antimicrobial resistance is one of largest threats to global health. On the
630 example of *Acinetobacter*, we have shown that the incorporation of a broad evolutionary
631 perspective can pinpoint individual genes or entire pathways that result in novel and viable
632 hypotheses of how the bacteria persist, feed off and interact with the host. At an early phase
633 of drug development, these candidates provide promising anchor points from which the
634 development of new therapeutic strategies to either prevent or treat *Acinetobacter*
635 infections can be initiated.
636

637 ## Methods

638 ### Data Acquisition

639 The full data set (Set-F) comprises all assemblies in the NCBI RefSeq data base (version 87)
640 stating 'Acinetobacter' in the 'organism' field [S1 Table]. From Set-F, we selected in total 232
641 representative strains covering all available type, reference and representative genomes, as
642 well as all validly named species (https://apps.szu.cz/anemec/Classification.pdf) for which a
643 genome sequence was available at the study onset are represented. We further picked
644 genomes of several *A. baumannii* strains that are of interest due to e.g., their context of
645 isolation as well as representatives of eight international clone types. Lastly, we included
646 genomes from Set-F into Set-R that allow for an increase of the fraction of total phylogenetic

647 diversity covered without compromising the quality (cf. S1 Text: Taxon Set Construction). An
648 overview of Set-R is provided in S2 Table.
649 To compile the *Proteobacteria* data set (n=1363), we selected all *Proteobacteria* represented
650 in NCBI RefSeq data base (version 204) and selected one representative per species, which
651 was either annotated as "reference" or "representative" strain.

652 ### International Clone Type (IC) and MLST assignments

653 For all strains in Set-F, we determined the sequence type with MLSTcheck v2.1.17 [113]
654 using two different schemes, Oxford [114] and Pasteur [115], that were obtained from the
655 PubMLST website (http://pubmlst.org/abaumannii/). All members of the *A. baumannii* clade
656 were assigned to an international clone (IC) type whenever we found literature evidence
657 that the predicted sequence types and IC were unambiguously linked using the following
658 publications as a source: [35, 115-124]. The final assignments are provided in S1 Table and
659 S2 Table based on strain typing results provided in S11 Table.

660 ### Phylogenetic diversity

661 Phylogenetic diversity scores of SET-F and SET-R were computed with PDA v.1.0.3 [125]
662 using the options *-k=234:3027* and *-if=handpicked.list* based on the ML tree of SET-F. 25 low
663 quality assemblies were pruned prior to this analysis to avoid overestimation because of
664 long branches resulting from sequencing errors. A detailed description of the quality
665 assessment is provided in S1 Text: Taxon Set Construction.

666 ### Average Nucleotide Identity

667 All genome sequences within Set-R were pair-wise aligned with Nucmer v3 [126], and the
668 Average Nucleotide Identities (ANIm) were calculated with the script
669 average_nucleotide_identity.py from the Python package pyani v0.2.7 [127] using the
670 following options*: -i ./genomes -o ./output/ -m ANIm -g --gmethod seaborn –maxmatch*.

671 ## Ortholog Search

672 All against all orthology searches were performed with OMA standalone v.2.2.0 [45] and
673 default parameter settings, except for decreasing the minimum length threshold for
674 sequences considered (MinLen) to 40 residues. Targeted ortholog searches were performed
675 with fDOG [128] using the OMA orthologous groups from SET-R to train the profile hidden
676 Markov models.

677 ## Inference of the pan and core gene sets (SET-R)

678 The pan-genome size was calculated as the sum of the number of OGs and number of strain
679 specific proteins (i.e. proteins without orthologs). For the rarefaction analysis, we identified
680 the number of new orthologs and singletons per added genome in Set-R. The core genome
681 was defined as the subset of OGs where each taxon of Set-R contributed exactly one
682 ortholog. However, the strict definition of the core genome, yielded very small core genome
683 estimates due to qualitative differences and incomplete draft genomes (S1 Text: Table 1).
684 We therefore relaxed the core genome definition and allowed a core-gene to be absent in
685 1% of the genomes (max. 3 out of the 232) in Set-R (see S1 Text for further details). To
686 obtain these values as a function of the number of genomes considered, we simulated a
687 sequential inclusion of genomes in SET-R. Following the approach from Tettelin et al. [129],
688 Pan- and core genome sizes were extrapolated by fitting the power law function $yP = \kappa Pn$
689 $^{\gamma} + c$ and an exponential decaying function $yc = \kappa c \exp[-n^{*}\tau c] + \Omega$, respectively, with
690 nonlinear least-squares (Levenberg–Marquardt algorithm). Given the large taxon set in our
691 study, we limited the simulation to 100 random permutations of a sequential inclusion.

692 ## Phylogenetic tree reconstruction

693 SET-R: Multiple sequence alignments (MSAs) for each orthologous group were generated
694 with MAFFT-LiNSI [130] (v7.394, default parameters). Next, each protein's CDS was obtained

695 and PAL2NAL v14 [131] with the option *-codontable 11* was employed to infer protein
696 sequence guided nucleotide MSAs. Supermatrices built from concatenated MSAs served as
697 input for the maximum likelihood (ML) tree reconstruction. The best fitting substitution
698 model (GTR + empirical base frequencies + free rate heterogeneity with 10 categories) was
699 determined using IQ-Tree v1.6.8 [132] using the option *-m TEST*, and the following
700 parameters were used for the ML tree reconstruction with IQ-Tree: *-m GTR+F+R10 -nt 6*
701 *AUTO -bb 1000 -alrt 1000*. Additionally, statistical branch supports were assessed with 1000
702 repetitions of UF bootstraps and SH-aLR branch tests. Trees were outgroup-rooted with
703 *Moraxella catarrhalis* (strains BBH18 and FDAARGOS_213) and *Alkanindiges illinoisensis* DSM
704 15370.
705 SET-F: The phylogenetic tree reconstruction of SET-F followed the same general work flow as
706 described for Set-R. However, to decrease the computational burden, trees were computed
707 on the amino acid sequence alignments. MAFFT was run with the '--auto' parameter.
708 Supermatrices of resulting protein MSAs were used as input for ML tree reconstruction with
709 IQ-TREE ('-alrt 1000 -bb 1000 -nt 8 -m LG+I+G+').
710 Majority-rule consensus trees were computed with SplitsTree v4.14.4 [133].

711 Inference of hierarchical orthologous groups and reconstruction of last common ancestor
712 dispensable and core gene sets

713 Hierarchical orthologous groups (HOGs) for the Set-R were inferred from the pairwise OMA
714 orthology relations and the consensus ML species tree using the GETHOG algorithm [134] as
715 implemented in OMA stand-alone v. 2.2.0. The Set-R pan genome was stratified by assigning
716 each HOG to the internal node of the tree that represents the last common ancestor (LCA) of
717 the two most distantly related taxa in the HOG. Clade-specific losses of a gene were inferred
718 when all members of a clade lacked an ortholog that was assigned to an evolutionarily older

719 node. On this basis, we reconstructed the pan-genomes for each internal node of the tree as
720 following: We united all genes assigned to internal nodes on the path from the root to the
721 node under study and removed the union of genes that have been lost on this path.

722 ## GO annotation and GO term enrichment analysis

723 All 502,095 unique protein accessions represented in SET-R were mapped to uniprotKB
724 identifiers (UniProt accessed 9th of Febr. 2018) to obtain the annotated gene ontology (GO)
725 terms [135]. Significantly enriched GO terms were identified using a two-tailed Fischer's
726 exact test at a significance level of 0.05. Multiple test correction was done by computing the
727 false discovery rate for each term and considering only terms with an FDR < 0.05. For GO
728 term enrichment analyses at the individual internal nodes of the tree, we used the LCA pan
729 gene sets as background sets (population) and the set of genes assigned to this node as test
730 (study) set. Both sets were limited to include only genes for which orthologs are represented
731 in at least 50% of the taxa descending from this node (analysis without this cutoff is included
732 in S5 Table 3). Information about the GO terms were pulled from
733 http://purl.obolibrary.org/obo/go/go-basic.obo (accessed 22nd of June 2018). Visualization
734 of tree maps were performed with REVIGO [136].

735 ## Prediction of Secreted Proteins

736 Unless the subcellular localization of a protein was provided by uniprot, we prediceted
737 Its subcellular localization with Psortb v3.06 [137] and ngLOC v1.0 [138]. All proteins
738 classified as 'Extracellular' or 'OuterMembrane' by either tool were combined into a set of
739 secreted and accessible proteins. Conflicting predictions or cases where Psortb labeled the
740 localization of a protein as 'unknown' were resolved in favor of the ngLOC classification as it
741 demonstrated higher precision for these classes when we benchmarked both against a test-
742 set of experimentally verified proteins published by Shen et al. [139].

743 ## Annotation of Protein Function

744 For each protein, we considered, where available, its functional annotation provided in
745 RefSeq, in the uniprot database, and the assigned GO terms. Additionally, we annotated the
746 proteins with KEGG Orthology (KO) identifiers [47] using GhostKoala v2.2 [140]. Eventually,
747 for proteins with reciprocal best blast hit orthologs in *A. baumannii* ATCC 17978, we
748 transferred the functional annotation provided by the COG database [46] (accessed in May
749 21st 2020).

750 ## Virulence Factor Identification

751 HOGs representing known virulence factors were identified via blastp searches (v2.10.1)
752 [141] against a custom database of virulence factors. To compile this database, we united
753 entries of PATRIC [21] and VFDB [48] (both accessed September 23rd 2020). Subsequently,
754 we clustered the proteins at 95% sequence similarity with cd-hit v4.6.4 [142] using the
755 options *-G -al 0.95 -s 0.9 -c 0.95* to reduce redundancy. A HOG represents a virulence factor
756 if any of the subsumed orthologs has a hit with >50% alignment coverage and an e-value
757 <0.01. For all such instance the best hit's annotation was transferred.

758 ## Identification of Evolutionarily Stable Gene Clusters (ESGC)

759 Using the genome of *A. baumannii* ATCC 19606 as a reference, we identified clusters of
760 consecutive genes with highly similar phyletic profiles. For this purpose, each protein-coding
761 gene's corresponding HOG and its profile were analyzed to obtain an 8-dimensional feature
762 vector comprised of the following values: (i-vi) clade-specific fractions of total taxa in *A.*
763 *baumannii* (B), *A. calcoaceticus* (CA), *A. haemolyticus* (HA), *A. baylyi* (BA), A. *lwoffii* (LW), *A.*
764 *brisouii* (BR), and *A. qingfengensis* (QI), (vii) the fraction of the total phylogenetic diversity of
765 the ACB clade covered, and (viii) the label of the inner node the HOG was assigned to (cf. Fig
766 3). Four genes were excluded from the orthology search due to their short length. For these,

767 we imputed the values using the mean values of the two flanking genes. Based on these
768 vector representations, we computed a pairwise dissimilarity matrix, using the Gower
769 dissimilarity index [143]. We then arranged the genes in a graph, where a vertex between a
770 gene and its downstream neighbor was drawn, if a) their pairwise dissimilarities is smaller
771 than the 5th percentile of the gene's dissimilarity distribution across the full gene set or b) if
772 the condition *a* is met by the two genes flanking the gene. The resulting set of connected
773 components were then extracted as candidate evolutionary units.

774 ### Candidate ESGC Abundance Statistics

775 For each gene in a candidate ESGC, we computed its retention difference (RD) as the
776 difference in the fraction of taxa within the ACB clade subtracted by the fraction of taxa
777 outside the ACB clade harboring an ortholog. The candidate ESGCs were then ranked by the
778 median of the RD across all genes in the cluster. As a further measure, we devised the
779 cluster-conservation score difference (CCD), which is calculated similar to the RD score, but
780 this time assessing presence of the cluster rather than that of an individual gene. A cluster
781 was considered present only if at least 80% of its genes (orthologs) were identified and at
782 least 25% of the gene order was conserved in the genome. Here, we treated the cluster as a
783 set of ordered and oriented (considering direction of transcription) two-element tuples.
784 Clusters with a CCD below 0.25 were not further considered. For each of the resulting top
785 150 ranked clusters, we inspected cluster conservation across the taxa in Set-R and Set-F
786 using Vicinator v0.32 (https://github.com/BIONF/Vicinator). Cluster boundaries of an ESGC
787 were manually curated, when the Vicinator analysis indicated a miss due to an individually
788 bloated abundance profile (false-positive orthologs or paralogs).

789 Kynurenine (Kyn) cluster deletion mutant and growth experiments with *A. baumannii* Ab04

790 Ab04 mutant with deletion of the locus of $ESGC_{ACB}$-0452 (Ab04 Δkyn) were constructed as
791 described previously [144]. Briefly, a FRT site-flanked apramycin resistance cassette was
792 amplified from a variant of pKD4 [145] with primers harboring 18-25 nucleotides matching in
793 sequence the flanking regions of the locus of $ESGC_{ACB}$-0452 (see S10 Table). Also, upstream
794 and downstream regions of $ESGC_{ACB}$-0452 were amplified and the obtained fragments were
795 assembled by overlap extension PCR. The PCR product was electroporated into *A. baumannii*
796 Ab04 electrocompetent cells carrying pAT04, which expresses the RecAB recombinase
797 induced with 2mM IPTG [1]. Mutants were selected with apramycin and integration of the
798 resistance marker was verified by PCR. To remove the resistance cassette, electrocompetent
799 mutants were transformed with pAT03 plasmid (which expresses the FLP recombinase) and
800 apramycin-sensitive clones of unmarked deletion mutants were obtained. Finally, the
801 mutant strains were confirmed by antibiotic resistance profile, PCR and genome sequencing.
802 Ab04 WT and Ab04 Δkyn mutant strains were grown in lysogeny broth (LB) liquid medium
803 under shaking conditions (200 rpm) at 37°C. Overnight cultures were washed three times
804 with PBS and diluted to an OD_{600} of 0.01 in 150 µl of M9 minimal medium supplemented
805 with 0.2% casamino acids (M9CAA), 2.5 mM L-kynurenine (Sigma, A8625) or 2.5 mM
806 anthranilic acid (Sigma, A89855) in 96-well plates, followed by incubation at 37°C under
807 shaking conditions in a BioTek microplate spectrophotometer. The OD_{600} values were
808 measured every 30 min for 16h. Three independent experiments were performed with three
809 wells per assay for each strain and condition.

810 Growth Experiments for Carnitine, Malate, and Glucarate

811 *A. baumannii, A. baylyi* and *A. calcoaceticus* strains were grown at 37 °C (*A. baumannii*) or
812 30 °C (*A. baylyi* and *A. calcoaceticus*) in mineral medium (MM) that consists of 50 mM

813 phosphate buffer, pH 6.8, and different salts (1 g NH_4Cl, 580 mg $MgSO_4 \times 7$ H2O, 100 mg

814 KNO_3, 67 mg $CaCl_2 \times 2\ H_2O$, 2 mg $(NH_4)_6Mo_7O_{24} \times 4\ H_2O$, 1 ml SL9 (per liter: 12.8 g Titriplex, 2

815 g $FeSO_4 \times 7\ H_2O$, 190 mg $CoCl_2 \times 6\ H_2O$, 122 mg $MnCl_2 \times 4\ H_2O$, 70 mg $ZnCl_2$, 36 mg $MoNa_2O_4$

816 $\times 2\ H_2O$, 24 mg $NiCl_2 \times 6\ H_2O$, 6 mg H_3BO_3, 2 mg $CuCl_2 \times H_2O$ per l medium; pH 6.5) [146] and

817 20 mM of the given carbon source. Precultures were grown in MM with 20 mM Na-acetate

818 as carbon source. Each value is the mean of +/- S.E.M. of at least three independent

819 measurements. Growth curves were fitted manually.

820

821

822

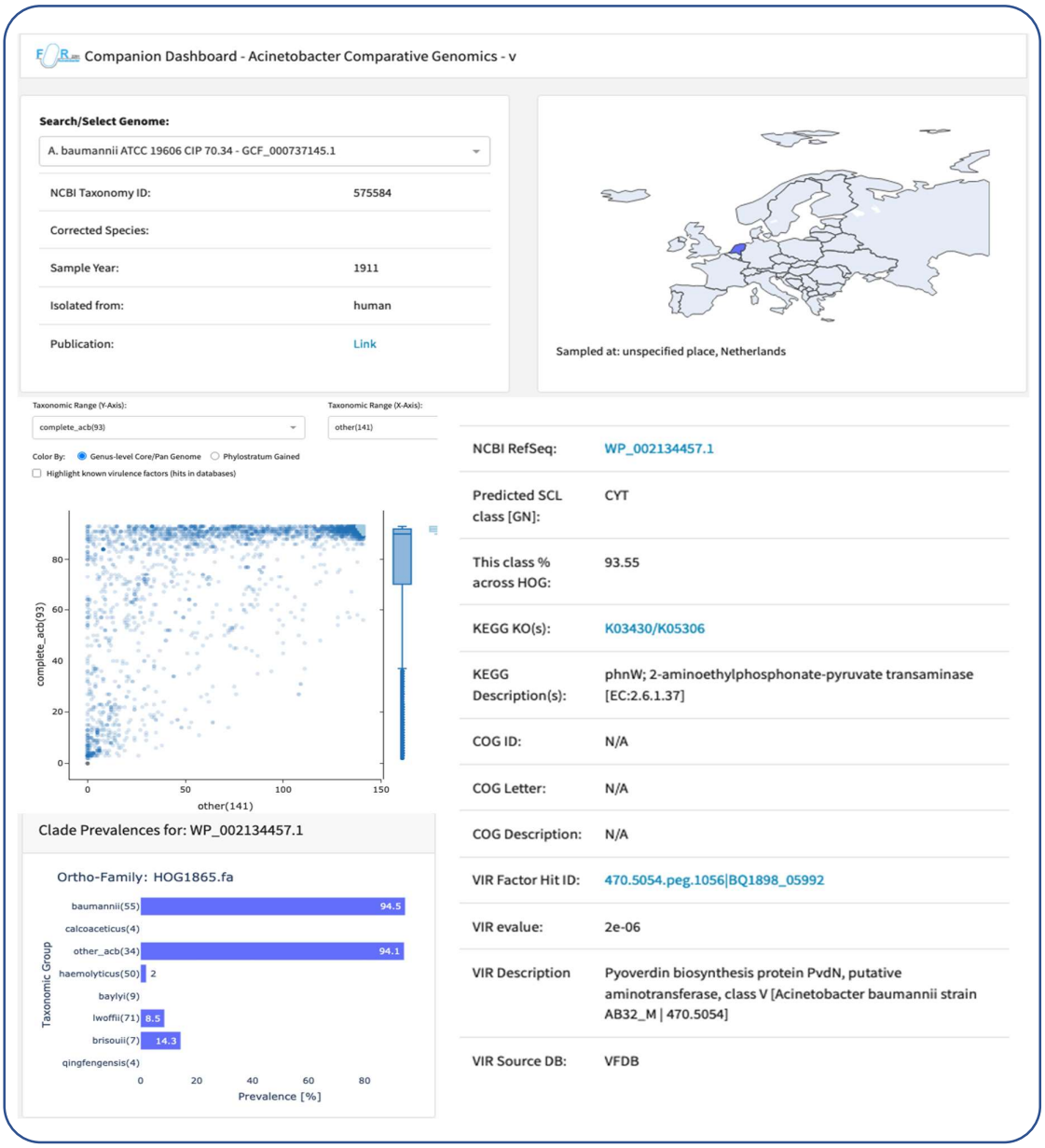


Fig 1. Aci-Dash – An interactive web-application to explore phylogenetic abundance patterns and accessory annotations for the Set-R pan-genome. *For each of the 232 strains, Aci-Dash provides further details about year and site of sampling. The interactive scatter plot reveals, for each gene of the selected strain, the abundance of orthologs in a user-defined in- and outgroup. For each gene individually, the ortholog abundance can be resolved on a clade level (cf. Fig 2), and further information including functional annotation transfer from KEGG and COG as well as known virulence factors is displayed.*

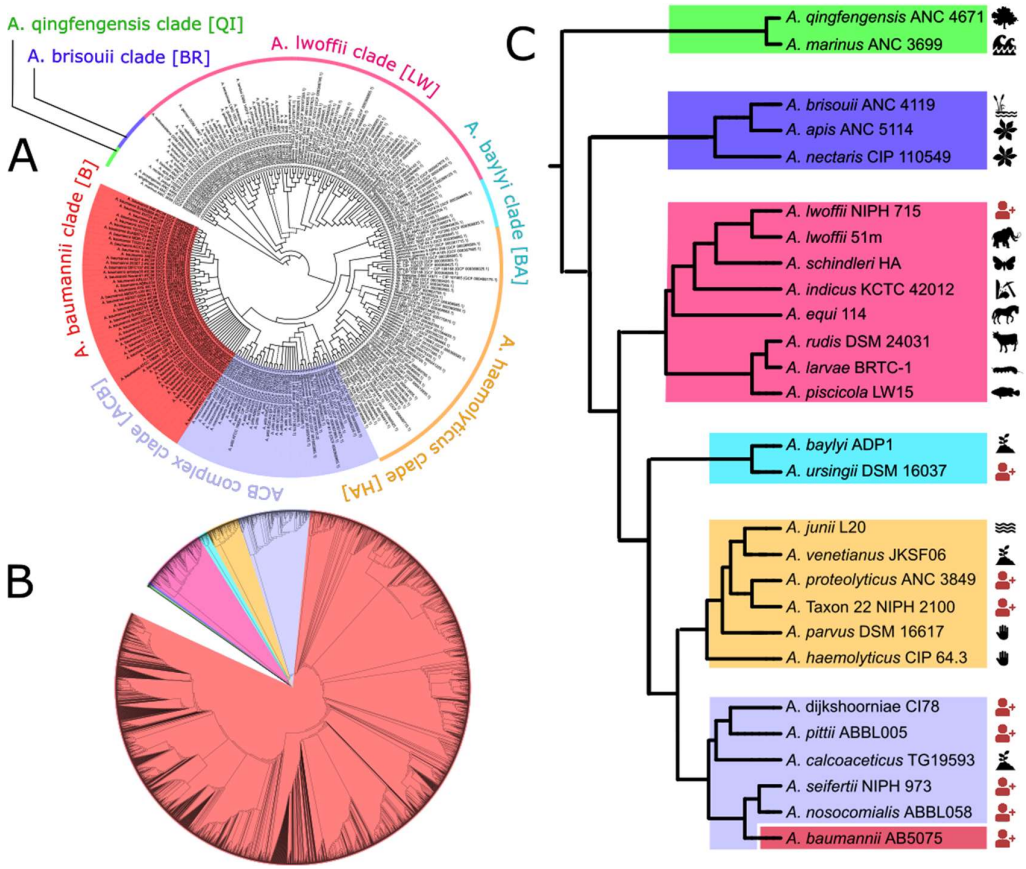


Fig 2. The phylogeny of the genus Acinetobacter. A) *Majority-rule consensus phylogeny of 232 Acinetobacter strains represented in SET-R. Solid branches are supported by all, and hatched branches by two out of three trees.* ***B)*** *The maximum likelihood tree for all 3052 taxa in Set-F. Colored clades represent the same clades as in A). A high-resolution image of this tree is provided in S8 Fig.* ***C)*** *The evolutionary backbone of the Acinetobacter genus with exemplarily chosen strains as clade representatives. The color scheme resembles that of Fig 2A. The pictograms next to the leaf labels indicate the sampling source of the particular strain. Red pictograms signal a strain that was isolated from an infected patient.*

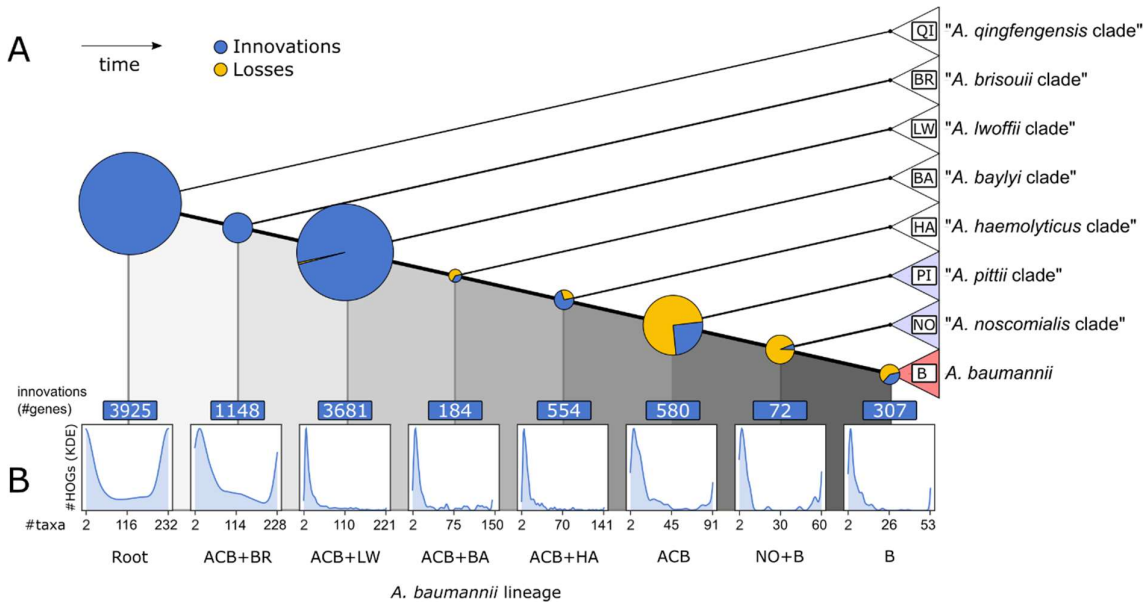
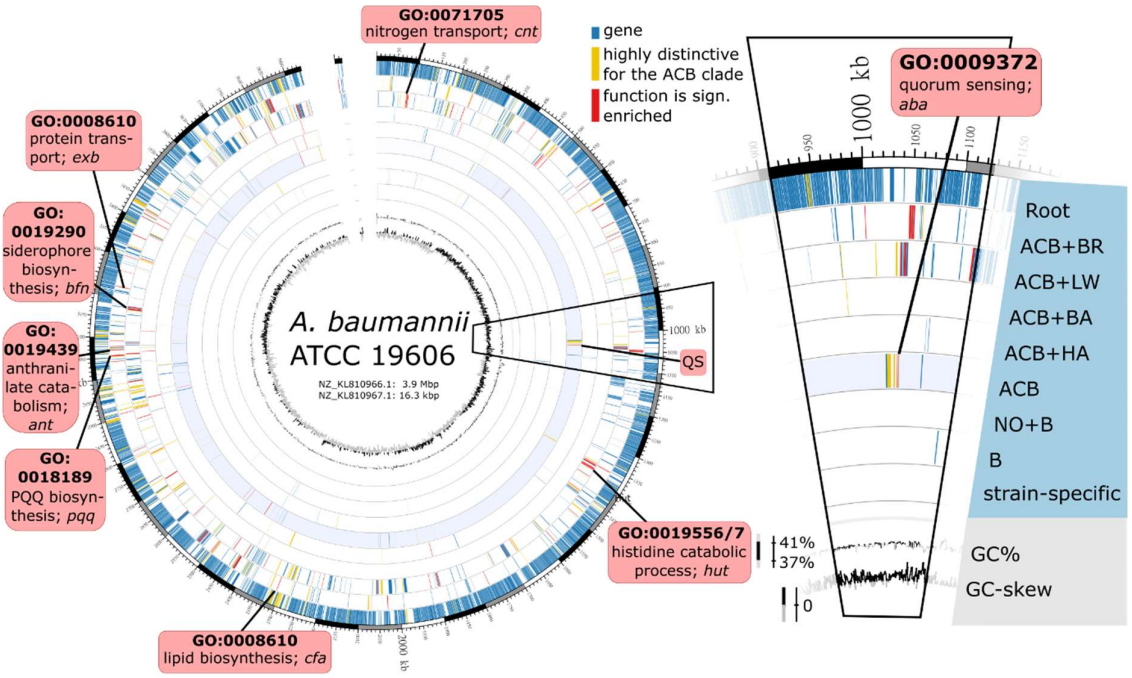


Fig 3. A phylogenetic stratification of the Acinetobacter pan-genome*. (A) Pie charts on the internal nodes indicate numbers of genes added (blue) or lost (yellow) from the respective pan-genome where the diameter is proportional to the sum of both gained and lost genes. Clades resemble those from Fig 2. (B) For each node, histograms represent the number of*

834 *descending taxa an added gene is represented in. The minimum value is 2, the maximum value is the number of taxa*
835 *subsumed in the corresponding clade.*

836



837
838 ***Fig 4. Phylogenetic strata in the protein-coding gene set of Ab. ATCC 19606.*** *Each protein-coding gene was assigned to one*
839 *of the nine layers specified in the inlay to the right (cf. Fig 3). All genes annotated with GO terms that were significantly*
840 *enriched in the individual layers are colored in red. Genes assigned to orthologous groups which were ranked within the 10*
841 *percent groups with highest retention differences (RD >=0.708, n=355) across the ATCC 19606 gene set are colored in*
842 *yellow. Red boxes highlight selected gene loci where neighboring genes contributed to the enrichment of the same GO term*
843 *(biological processes).*

844

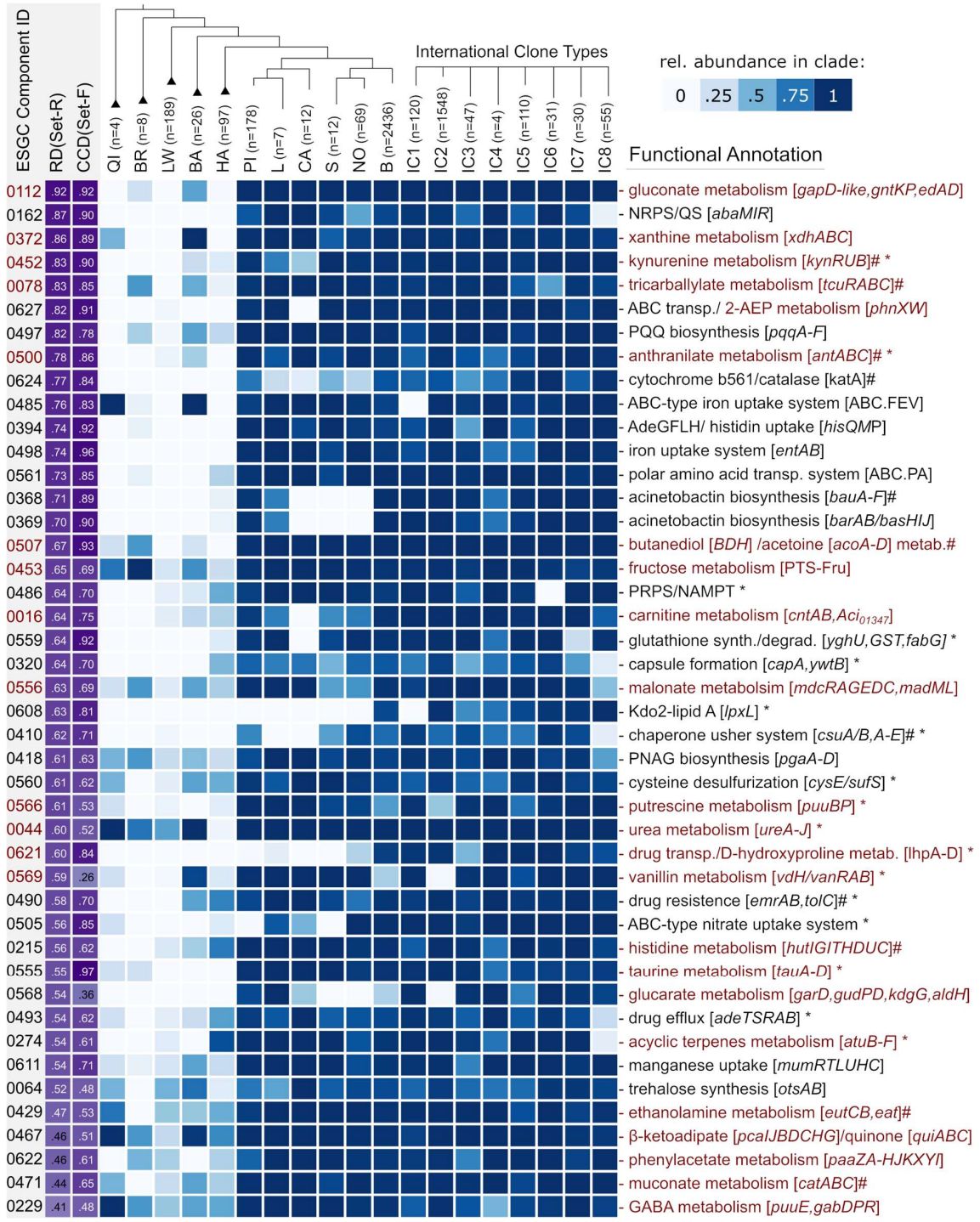


Fig 5. Phylogenetic profiles of 44 functionally annotated ESGC$_{ACB}$ across the Set-F. *Shown are the selection of ESGC$_{ACB}$ with the highest abundance difference between the ACB clade and non-ACB taxa (RD). Ids refer to the corresponding connected components provided in S6 Table 3. Gene clusters are ranked by decreasing RD. The CCD (cluster conservation difference, violet) is calculated as the difference between the relative fractions of ACB-genomes minus non-ACB genomes where the cluster is present. The heat map informs about the fraction of taxa per clade harboring the ESGC$_{ACB}$ ranging from 0 (white) to 1 (dark blue). The total number of subsumed taxa per clade are given next to the leaf label in the above tree. IC1-8 represent the cluster abundance in the 8 international clone types of A. baumannii. Clusters associated with metabolic pathways are highlighted in red. "#" marks an ESGC where we curated cluster boundaries based on literature evidence and confirmed cluster conservation using microsynteny (cf. Fig 4B). '*' marks ESGC$_{ACB}$ that encompass additional genes not covered by the functional annotation. Abbreviations: NRPS = non-ribosomal peptide synthesis; QS = quorum sensing; 2-AEP =2-aminoethylphosphonate; PNAG = polymeric β-1,6-linked N-acetylglucosamine; NAMPT = nicotinamide phosphoribosyl transferase; PRPS = 5-phosphoribosyl pyrophosphate synthetase; GABA = gamma-aminobutyric acid; PQQ = pyrroloquinoline quinone.*

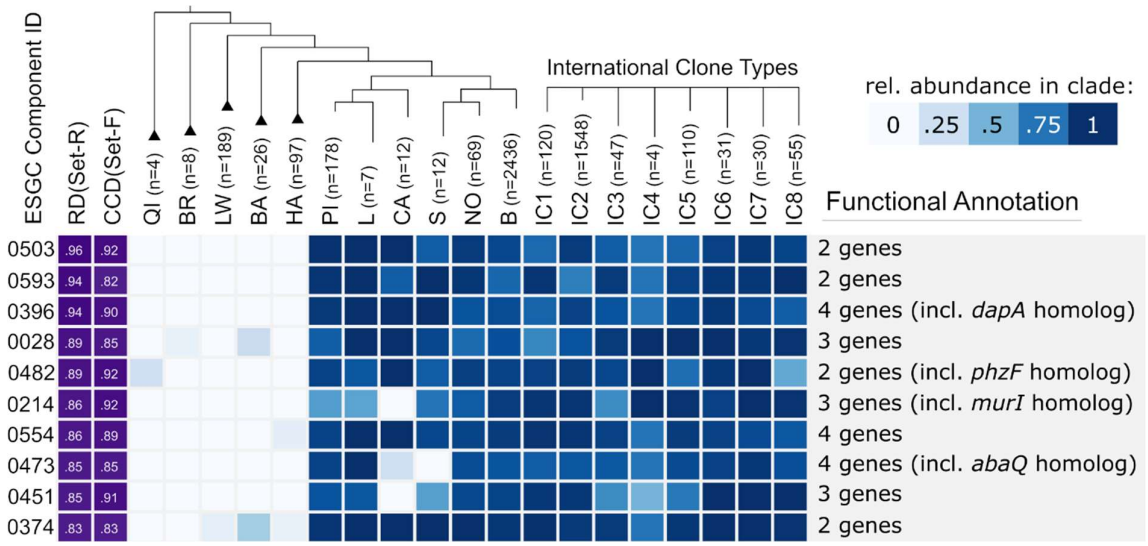


Fig 6. Top 10 ESGC$_{ACB}$ with unknown cluster function. *The genes of these clusters are mostly annotated as 'hypothetical" or "DUF-containing protein", i.e. proteins with annotated with a domain of unknown function. Due to their stable microsynteny and high prevalence across the ACB clade, they are highly interesting candidates for further functional characterization. The figure layout follows Fig 5.*

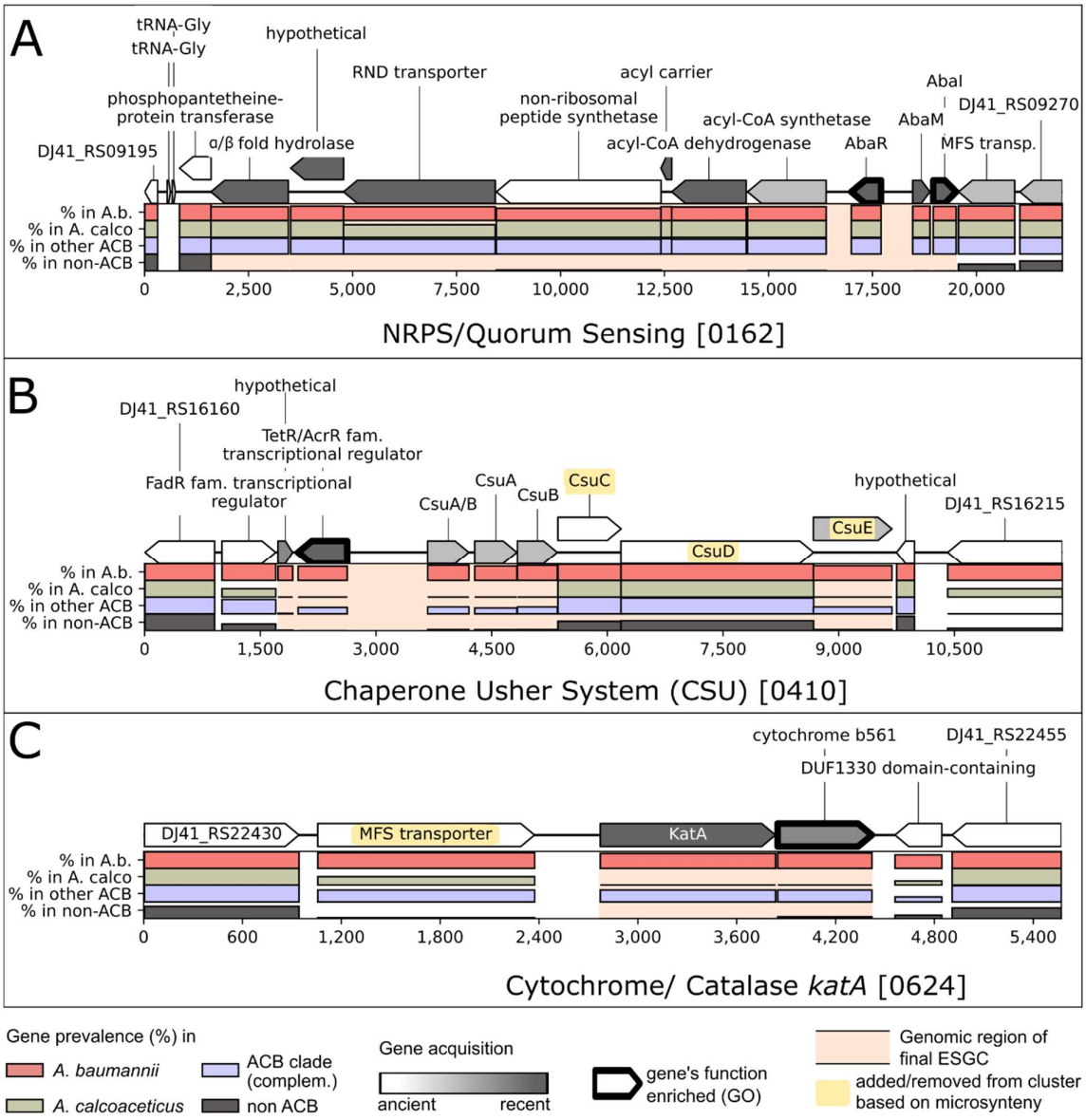


Fig 7. Three examples for $ESGC_{ACB}$ in the genome of Ab ATCC 19606. *Bar plots indicate relative abundance within the selected taxonomic groups (see legend). The $ESGC_{ACB}$ are embedded into two flanking protein-coding genes on each side that are not part of the cluster. Cluster boundaries are indicated by a pink background. Genes indicated in yellow have been either added to or excluded from the automatically generated clusters based on microsynteny analyses across Set-R. A) $ESGC_{ACB}$ – 0162 unites AbaR and AbaI (quorum sensing) with a non-ribosomal peptide synthetase cluster upstream. B) $ESGC_{ACB}$ – 0410 encompasses the cluster necessary for the Csu pilus formation. Note the deviating abundance pattern for CsuC and CsuD, which is due to the presence of prpC and prpD, two paralogous genes from the photoregulated pilus ABCD (prpABCD), in the respective orthologous groups. Microsynteny analyses confirmed that the entire Csu cluster forms one evolutionary unit (S1 Data id:0410). C) $ESGC_{ACB}$ – 0624 harbors a catalase, a cytochrome b561. Although the MFS transporter shares a similar abundance pattern across the taxa of Set-R, this transporter is not evolutionarily stably linked to the other two genes (S1 Data id:0624).*

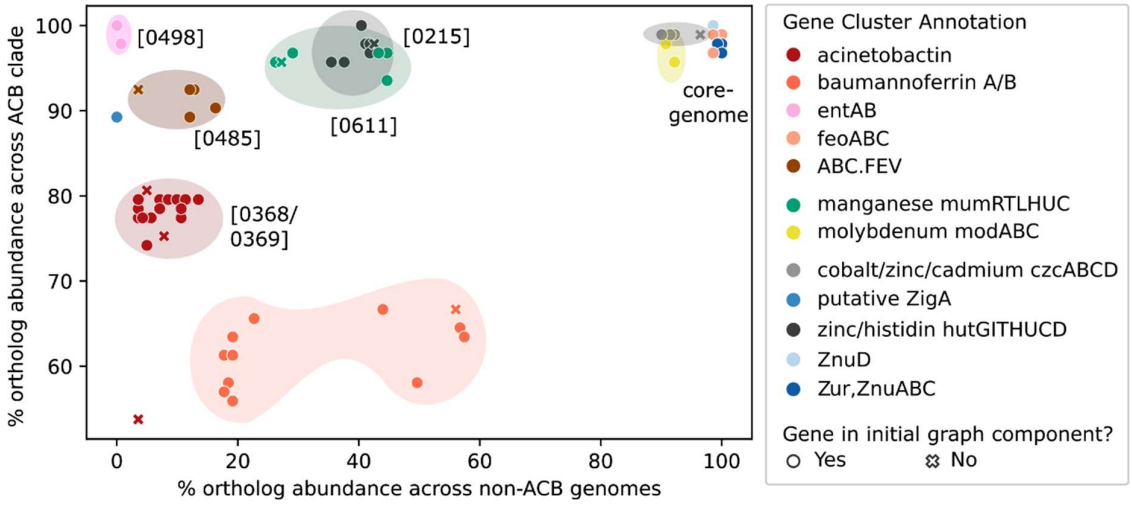


Fig 8. Abundances of micronutrient acquisition genes/clusters of Ab ATCC 19606 within and outside the ACB clade. *Absolute abundances were based on cluster-evaluation using microsynteny. Genes clusters are annotated with the $ESGC_{ACB}$ identifiers in Fig 5. Genes that are part of a cluster but missed by the initial ESGC compilation share the same color but are marked as x. Gene clusters at the top right, e.g., feoABC, represent micronutrient acquisition clusters belonging to the genus' core genome. Gene clusters at the top left, e.g., entAB, are confined to the ACB clade where they are ubiquitously present.*

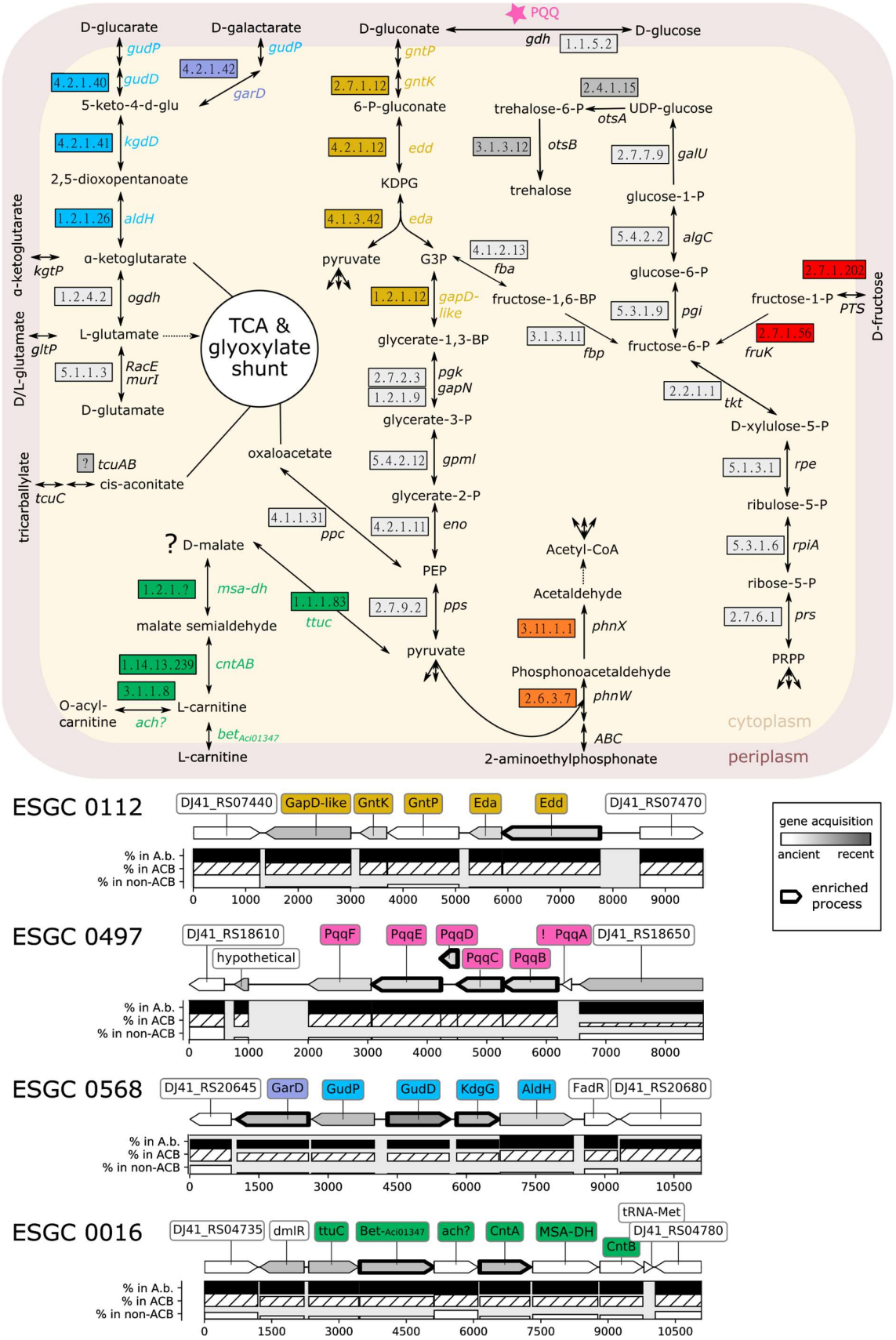


Fig 9. Members of the ACB clade have extended their basal carbohydrate metabolism. *The pathway map shows a model integrating four metabolic pathways represented by the ESGC$_{ACB}$ 0112 (Entner-Doudoroff pathway; yellow), 0497 (biosynthesis of Pyrroloquinoline quinone; magenta), 0568 (glucarate/galactarate catabolism; blue/violet) and 0016*

(carnitine catabolism; green) into the bacterial carbohydrate metabolism (grey boxes). Abbreviations: KDPG - 2-keto-3-deoxy-6-phosphogluconate; G3P – glycerinealdehyde-3-phosphate. The corresponding gene clusters are shown below the pathway map with one protein-coding flanking gene on either side that is not part of the cluster. The layout follows Fig 7. $ESGC_{ACB}$ 0112: GapD-like – glyceraldehyde-3-phosphate dehydrogenase [1.2.1.12]; GntK - gluconokinase [EC:2.7.1.12]; Edd - phosphogluconate dehydratase [EC:4.2.1.12]); Eda - 2-dehydro-3-deoxyphosphogluconate aldolase / (4S)-4-hydroxy-2-oxoglutarate aldolase [EC:4.1.2.14 4.1.3.42]; GntP - gluconate permease [E2.7.1.12]. Note, GntP shares high sequence similarity with a gluconate transporter (H+ symporter) in Escherichia coli (98% coverage, 45% identity). $ESGC_{ACB}$ 0497: pqqA-F. No abundance profile is shown for pqqA, since its length excluded it from orthology prediction (indicated by an '!'. See main text for details). $ESGC_{ACB}$ 0568: GarD - galactarate dehydratase [EC:4.2.1.42]; gudD - glucarate dehydratase [EC:4.2.1.40]; kdgD - 5-dehydro-4-deoxyglucarate dehydratase [EC:4.2.1.41]; aldH - 2,5-dioxopentanoate dehydrogenase [EC:1.2.1.26]; gudP - MFS transporter, D-glucarate/D-galactarate permease; FadR - Fatty acid metabolism regulator protein. $ESGC_{ACB}$ 0016: ttuc - D-malate dehydrogenase [EC:1.1.1.83]; Bet-Aci01347 - glycine/betaine transporter Aci01347; ach? – putative acylcarnitine hydrolase [EC:3.1.1.8]; CntAB – carnitine monooxygenase reductase subunit A and B [EC:1.14.13.239]; MSA-DH – malic semialdehyde dehydrogenase [EC:1.2.1.?]. Further colored pathways represent the clusters $ESGC_{ACB}$ 0627 (2-aminoethylphosphonate metabolism, orange), $ESGC_{ACB}$ 0453 (fructose transport/metabolism, red), $ESGC_{ACB}$ 0064 (trehalose biosynthesis, dark grey), and $ESGC_{ACB}$ 0078 tricarballylate metabolism, brown). Their corresponding genomic regions are available in S5 Fig.

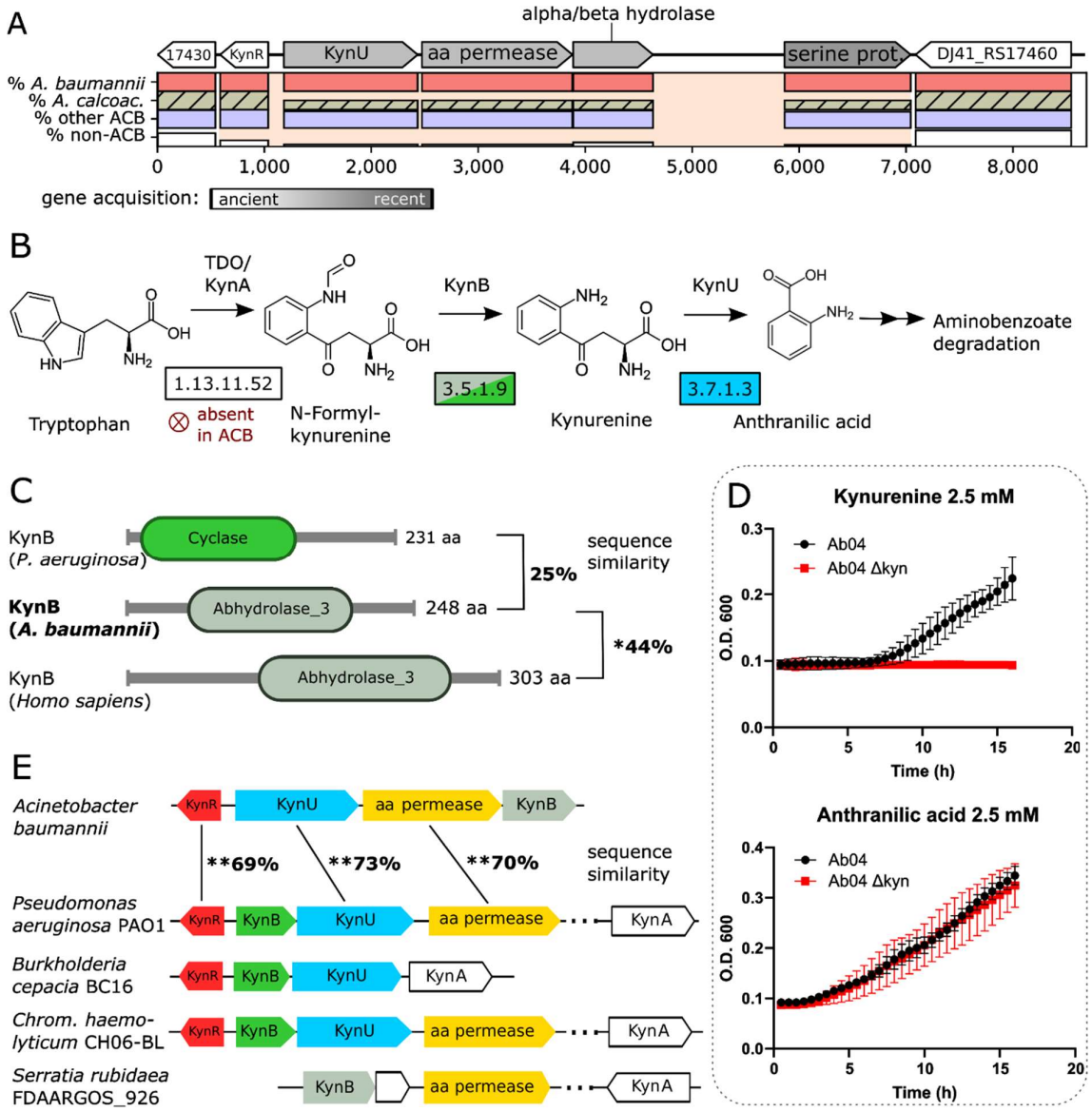


Fig 10. Evolutionary and functional characterization of the A. baumannii kynurenine pathway. *A) The locus of ESGC$_{ACB}$ 0452 in A. baumannii ATCC 19606. KynR - Lrp/AsnC family transcriptional regulator; KynU - kynurenine hydrolase; amino acid (aa) permease (uncharacterized); alpha/beta hydrolase (uncharacterized; putative kynB); serine proteinase. Cluster layout follows that of Fig 7. B) The Kynurenine pathway of tryptophan degradation. EC numbers of the enzymes catalyzing the individual reactions are given in the boxes. KynB is represented in two versions in bacteria (see C), which are represented by different colors. C) The alpha/beta hydrolase shares a significant sequence similarity and the presence of the Abhydrolase_3 Pfam domain (PF04199) with the human kynurenine formamidase (KynB or KFA) but not with KynB of P. aeruginosa, which instead harbors a Cyclase Pfam domain (PF04199) suggesting an independent evolutionary origin. D) The kynurenine cluster is necessary and sufficient for growth on kynurenine but not on anthranilic acid. Growth of Ab 04 and Ab 04 Δkyn on minimal medium supplemented with kynurenine (top) and anthranilate (bottom). E) Phylogenetic profile of the A. baumannii Kyn cluster. Numbers between the corresponding genes in Ab ATCC 19606 and PA01 represent percent sequence similarity on the amino acid level. KynB of Ab ATCC 19606 and PAO1 are not significantly similar. The Kyn cluster is shared among many proteobacterial (opportunistic) pathogens where the the Pseudomonas-type KynB is prevalent. The human-type KynB has homologs e.g., in Serratia (including S. rubideae). KynA is almost always encoded at a distant locus and entirely absent in Acinetobacter.*

927 ## Declarations

928 **Availability of data and materials**

929 Genomic data analyzed during the current study are publicly available in the NCBI Reference
930 Sequence Database (RefSeq) version 87 and 204. For each selected genome, the assembly
931 accession listed in S1 Table and S9 Table, respectively. From individual entries, we obtained
932 genome annotation files (*.gff and *_feature_table.txt), protein sequences and coding
933 sequences (*.faa and *cds_from_genomic.fna), and genome sequence (*_genomic.fna) files
934 from the database [ftp://ftp.ncbi.nlm.nih.gov/refseq/release/release-catalog/archive/].
935 In addition, the datasets used and/or analyzed during the current study are available from
936 the corresponding author on reasonable request.
937 All data generated or analyzed during this study and necessary for the interpretation are
938 included in this published article as supplementary information.
939 Our custom tool Vicinator is linked (https://github.com/BIONF/Vicinator) and publicly
940 available under the GNP license 3.0.
941

942 **Competing interests**

943 The authors declare that they have no competing interests.
944

945 **Funding**

946 This study was supported by a grant by the German Research Foundation (DFG) in the scope
947 of the Research Group FOR2251 “Adaptation and persistence of A. baumannii.” Grant ID EB-
948 285-2/2 to IE, AV 9/7-2 to BA, GO 2491/1-2 to SG, and WI 3272/3-2 to GW. MFF was
949 supported by grants from the National Institute of Allergy and Infectious Diseases (grant

R01AI144120). Cloud computational resources through the de.NBI cloud are granted by the German Bundesministerium für Bildung und Forschung grant FKZ 031A533B to AG.

Authors' contribution

BD and IE conceived the study. BD performed data collection and curation, tool and app development and performed bioinformatics/statistical analyses. BD and IE together with GDV, MF, GW, BA and SG interpreted the data. SD and GDV performed and interpreted wild type and mutant strain growth experiments on kynurenine and anthranilate. JB performed growth experiments on carnitine, malate and gluconate. MAD developed the CI/CD pipeline for a containerized deployment of Aci-Dash in a cloud-platform under the supervision of AG. BD and IE drafted the manuscript, and all authors contributed to the final manuscript. All authors read and approved the final manuscript.

Acknowledgements

The authors wish to thank all researchers for making annotated genome sequences available to the public domain, Prof. Alexandr Nemec for providing on his website an accessible and up-to-date resource of the current species nomenclature of *Acinetobacter*, Vinh Tran for support with fDOG, and Ruben Iruegas for helpful discussion.

References

1. Garnacho-Montero J, Amaya-Villar R. Multiresistant Acinetobacter baumannii infections: epidemiology and management. Curr Opin Infect Dis. 2010;23(4):332-9. Epub 2010/06/29. doi: 10.1097/QCO.0b013e32833ae38b. PubMed PMID: 20581674.
2. Peleg AY, Seifert H, Paterson DL. Acinetobacter baumannii: emergence of a successful pathogen. Clin Microbiol Rev. 2008;21(3):538-82. Epub 2008/07/16. doi: 10.1128/CMR.00058-07. PubMed PMID: 18625687; PubMed Central PMCID: PMCPMC2493088.
3. Gupta N, Gandham N, Jadhav S, Mishra RN. Isolation and identification of Acinetobacter species with special reference to antibiotic resistance. J Nat Sci Biol Med.

979 2015;6(1):159-62. Epub 2015/03/27. doi: 10.4103/0976-9668.149116. PubMed PMID:
980 25810655; PubMed Central PMCID: PMCPMC4367029.
981 4. Sahl JW, Gillece JD, Schupp JM, Waddell VG, Driebe EM, Engelthaler DM, et al.
982 Evolution of a pathogen: a comparative genomics analysis identifies a genetic pathway to
983 pathogenesis in Acinetobacter. PLoS ONE. 2013;8(1):e54287. Epub 2013/02/01. doi:
984 10.1371/journal.pone.0054287. PubMed PMID: 23365658; PubMed Central PMCID:
985 PMC3554770.
986 5. Visca P, Seifert H, Towner KJ. Acinetobacter infection--an emerging threat to human
987 health. Iubmb Life. 2011;63(12):1048-54. Epub 2011/10/19. doi: 10.1002/iub.534. PubMed
988 PMID: 22006724.
989 6. Wong D, Nielsen TB, Bonomo RA, Pantapalangkoor P, Luna B, Spellberg B. Clinical and
990 Pathophysiological Overview of Acinetobacter Infections: a Century of Challenges. Clin
991 Microbiol Rev. 2017;30(1):409-47. Epub 2016/12/16. doi: 10.1128/CMR.00058-16. PubMed
992 PMID: 27974412; PubMed Central PMCID: PMCPMC5217799.
993 7. Weber BS, Harding CM, Feldman MF. Pathogenic Acinetobacter: from the Cell
994 Surface to Infinity and Beyond. J Bacteriol. 2015;198(6):880-7. Epub 2015/12/30. doi:
995 10.1128/JB.00906-15. PubMed PMID: 26712938; PubMed Central PMCID:
996 PMCPMC4772598.
997 8. Lob SH, Hoban DJ, Sahm DF, Badal RE. Regional differences and trends in
998 antimicrobial susceptibility of Acinetobacter baumannii. Int J Antimicrob Agents.
999 2016;47(4):317-23. Epub 2016/03/30. doi: 10.1016/j.ijantimicag.2016.01.015. PubMed
1000 PMID: 27020541.
1001 9. Touchon M, Cury J, Yoon EJ, Krizova L, Cerqueira GC, Murphy C, et al. The genomic
1002 diversification of the whole Acinetobacter genus: origins, mechanisms, and consequences.
1003 Genome Biol Evol. 2014;6(10):2866-82. Epub 2014/10/15. doi: 10.1093/gbe/evu225.
1004 PubMed PMID: 25313016; PubMed Central PMCID: PMCPMC4224351.
1005 10. Giammanco A, Cala C, Fasciana T, Dowzicky MJ. Global Assessment of the Activity of
1006 Tigecycline against Multidrug-Resistant Gram-Negative Pathogens between 2004 and 2014
1007 as Part of the Tigecycline Evaluation and Surveillance Trial. mSphere. 2017;2(1). Epub
1008 2017/01/27. doi: 10.1128/mSphere.00310-16. PubMed PMID: 28124025; PubMed Central
1009 PMCID: PMCPMC5244261.
1010 11. Wilharm G, Piesker J, Laue M, Skiebe E. DNA uptake by the nosocomial pathogen
1011 Acinetobacter baumannii occurs during movement along wet surfaces. J Bacteriol.
1012 2013;195(18):4146-53. Epub 2013/07/16. doi: 10.1128/JB.00754-13. PubMed PMID:
1013 23852865; PubMed Central PMCID: PMCPMC3754747.
1014 12. Gallagher LA, Ramage E, Weiss EJ, Radey M, Hayden HS, Held KG, et al. Resources for
1015 Genetic and Genomic Analysis of Emerging Pathogen Acinetobacter baumannii. J Bacteriol.
1016 2015;197(12):2027-35. Epub 2015/04/08. doi: 10.1128/JB.00131-15. PubMed PMID:
1017 25845845; PubMed Central PMCID: PMCPMC4438207.
1018 13. Weidensdorfer M, Ishikawa M, Hori K, Linke D, Djahanschiri B, Iruegas R, et al. The
1019 Acinetobacter trimeric autotransporter adhesin Ata controls key virulence traits of
1020 Acinetobacter baumannii. Virulence. 2019;10(1):68-81. Epub 2019/12/25. doi:
1021 10.1080/21505594.2018.1558693. PubMed PMID: 31874074; PubMed Central PMCID:
1022 PMCPMC6363060.
1023 14. Wieland K, Chhatwal P, Vonberg RP. Nosocomial outbreaks caused by Acinetobacter
1024 baumannii and Pseudomonas aeruginosa: Results of a systematic review. Am J Infect
1025 Control. 2018;46(6):643-8. Epub 2018/02/06. doi: 10.1016/j.ajic.2017.12.014. PubMed
1026 PMID: 29398072.

1027 15. Lowman W, Kalk T, Menezes CN, John MA, Grobusch MP. A case of community-
1028 acquired Acinetobacter baumannii meningitis - has the threat moved beyond the hospital? J
1029 Med Microbiol. 2008;57(Pt 5):676-8. Epub 2008/04/26. doi: 10.1099/jmm.0.47781-0.
1030 PubMed PMID: 18436607.
1031 16. Banin E, Hughes D, Kuipers OP. Editorial: Bacterial pathogens, antibiotics and
1032 antibiotic resistance. FEMS Microbiol Rev. 2017;41(3):450-2. Epub 2017/05/10. doi:
1033 10.1093/femsre/fux016. PubMed PMID: 28486583.
1034 17. Tacconelli E, Carrara E, Savoldi A, Harbarth S, Mendelson M, Monnet DL, et al.
1035 Discovery, research, and development of new antibiotics: the WHO priority list of antibiotic-
1036 resistant bacteria and tuberculosis. Lancet Infect Dis. 2018;18(3):318-27. Epub 2017/12/26.
1037 doi: 10.1016/S1473-3099(17)30753-3. PubMed PMID: 29276051.
1038 18. Garcia-Quintanilla M, Pulido MR, Lopez-Rojas R, Pachon J, McConnell MJ. Emerging
1039 therapies for multidrug resistant Acinetobacter baumannii. Trends Microbiol.
1040 2013;21(3):157-63. doi: 10.1016/j.tim.2012.12.002. PubMed PMID: 23317680.
1041 19. Morris FC, Dexter C, Kostoulias X, Uddin MI, Peleg AY. The Mechanisms of Disease
1042 Caused by Acinetobacter baumannii. Frontiers in microbiology. 2019;10:1601. Epub
1043 2019/08/06. doi: 10.3389/fmicb.2019.01601. PubMed PMID: 31379771; PubMed Central
1044 PMCID: PMCPMC6650576.
1045 20. Sarshar M, Behzadi P, Scribano D, Palamara AT, Ambrosi C. Acinetobacter baumannii:
1046 An Ancient Commensal with Weapons of a Pathogen. Pathogens. 2021;10(4). Epub
1047 2021/04/04. doi: 10.3390/pathogens10040387. PubMed PMID: 33804894; PubMed Central
1048 PMCID: PMCPMC8063835.
1049 21. Davis JJ, Wattam AR, Aziz RK, Brettin T, Butler R, Butler RM, et al. The PATRIC
1050 Bioinformatics Resource Center: expanding data and analysis capabilities. Nucleic Acids Res.
1051 2020;48(D1):D606-D12. Epub 2019/11/02. doi: 10.1093/nar/gkz943. PubMed PMID:
1052 31667520; PubMed Central PMCID: PMCPMC7145515.
1053 22. Liu B, Zheng D, Jin Q, Chen L, Yang J. VFDB 2019: a comparative pathogenomic
1054 platform with an interactive web interface. Nucleic Acids Res. 2019;47(D1):D687-D92. Epub
1055 2018/11/06. doi: 10.1093/nar/gky1080. PubMed PMID: 30395255; PubMed Central PMCID:
1056 PMCPMC6324032.
1057 23. Juttukonda LJ, Green ER, Lonergan ZR, Heffern MC, Chang CJ, Skaar EP. Acinetobacter
1058 baumannii OxyR Regulates the Transcriptional Response to Hydrogen Peroxide. Infect
1059 Immun. 2019;87(1). Epub 2018/10/10. doi: 10.1128/IAI.00413-18. PubMed PMID: 30297527;
1060 PubMed Central PMCID: PMCPMC6300632.
1061 24. McConnell MJ, Actis L, Pachon J. Acinetobacter baumannii: human infections, factors
1062 contributing to pathogenesis and animal models. FEMS Microbiol Rev. 2013;37(2):130-55.
1063 Epub 2012/05/10. doi: 10.1111/j.1574-6976.2012.00344.x. PubMed PMID: 22568581.
1064 25. Harding CM, Hennon SW, Feldman MF. Uncovering the mechanisms of Acinetobacter
1065 baumannii virulence. Nat Rev Microbiol. 2018;16(2):91-102. Epub 2017/12/19. doi:
1066 10.1038/nrmicro.2017.148. PubMed PMID: 29249812; PubMed Central PMCID:
1067 PMCPMC6571207.
1068 26. Giles SK, Stroeher UH, Eijkelkamp BA, Brown MH. Identification of genes essential for
1069 pellicle formation in Acinetobacter baumannii. BMC Microbiol. 2015;15:116. Epub
1070 2015/06/07. doi: 10.1186/s12866-015-0440-6. PubMed PMID: 26047954; PubMed Central
1071 PMCID: PMCPMC4457973.
1072 27. Green ER, Juttukonda LJ, Skaar EP. The Manganese-Responsive Transcriptional
1073 Regulator MumR Protects Acinetobacter baumannii from Oxidative Stress. Infect Immun.

1074 2020;88(3). Epub 2019/12/04. doi: 10.1128/IAI.00762-19. PubMed PMID: 31792075;
1075 PubMed Central PMCID: PMCPMC7035938.
1076 28. Ramirez MS, Penwell WF, Traglia GM, Zimbler DL, Gaddy JA, Nikolaidis N, et al.
1077 Identification of Potential Virulence Factors in the Model Strain Acinetobacter baumannii
1078 A118. Frontiers in microbiology. 2019;10:1599. Epub 2019/08/10. doi:
1079 10.3389/fmicb.2019.01599. PubMed PMID: 31396168; PubMed Central PMCID:
1080 PMCPMC6663985.
1081 29. Eisenreich W, Dandekar T, Heesemann J, Goebel W. Carbon metabolism of
1082 intracellular bacterial pathogens and possible links to virulence. Nat Rev Microbiol.
1083 2010;8(6):401-12. Epub 2010/05/11. doi: 10.1038/nrmicro2351. PubMed PMID: 20453875.
1084 30. Passalacqua KD, Charbonneau ME, O'Riordan MXD. Bacterial Metabolism Shapes the
1085 Host-Pathogen Interface. Microbiol Spectr. 2016;4(3). Epub 2016/06/24. doi:
1086 10.1128/microbiolspec.VMBF-0027-2015. PubMed PMID: 27337445; PubMed Central
1087 PMCID: PMCPMC4922512.
1088 31. Nemec A. Classification and nomenclature of species in the genus Acinetobacter
1089 2020. Available from: https://apps.szu.cz/anemec/Classification.pdf.
1090 32. Al Atrouni A, Joly-Guillou ML, Hamze M, Kempf M. Reservoirs of Non-baumannii
1091 Acinetobacter Species. Front Microbiol. 2016;7:49. Epub 2016/02/13. doi:
1092 10.3389/fmicb.2016.00049. PubMed PMID: 26870013; PubMed Central PMCID:
1093 PMCPMC4740782.
1094 33. Wallace L, Daugherty SC, Nagaraj S, Johnson JK, Harris AD, Rasko DA. Use of
1095 Comparative Genomics To Characterize the Diversity of Acinetobacter baumannii
1096 Surveillance Isolates in a Health Care Institution. Antimicrob Agents Chemother.
1097 2016;60(10):5933-41. Epub 2016/07/28. doi: 10.1128/AAC.00477-16. PubMed PMID:
1098 27458211; PubMed Central PMCID: PMCPMC5038335.
1099 34. Holt K, Kenyon JJ, Hamidian M, Schultz MB, Pickard DJ, Dougan G, et al. Five decades
1100 of genome evolution in the globally distributed, extensively antibiotic-resistant
1101 Acinetobacter baumannii global clone 1. Microb Genom. 2016;2(2):e000052. Epub
1102 2017/03/30. doi: 10.1099/mgen.0.000052. PubMed PMID: 28348844; PubMed Central
1103 PMCID: PMCPMC5320584.
1104 35. Chan AP, Sutton G, DePew J, Krishnakumar R, Choi Y, Huang XZ, et al. A novel method
1105 of consensus pan-chromosome assembly and large-scale comparative analysis reveal the
1106 highly flexible pan-genome of Acinetobacter baumannii. Genome Biol. 2015;16:143. Epub
1107 2015/07/22. doi: 10.1186/s13059-015-0701-6. PubMed PMID: 26195261; PubMed Central
1108 PMCID: PMCPMC4507327.
1109 36. Imperi F, Antunes LCS, Blom J, Villa L, Iacono M, Visca P, et al. The Genomics of
1110 Acinetobacter baumannii: Insights into Genome Plasticity, Antimicrobial Resistance and
1111 Pathogenicity. Iubmb Life. 2011;63(12):1068-74. doi: Doi 10.1002/Iub.531. PubMed PMID:
1112 ISI:000297947900005.
1113 37. Abbott I, Cerqueira GM, Bhuiyan S, Peleg AY. Carbapenem resistance in
1114 Acinetobacter baumannii: laboratory challenges, mechanistic insights and therapeutic
1115 strategies. Expert Rev Anti Infect Ther. 2013;11(4):395-409. Epub 2013/04/10. doi:
1116 10.1586/eri.13.21. PubMed PMID: 23566149.
1117 38. Eijkelkamp BA, Stroeher UH, Hassan KA, Paulsen IT, Brown MH. Comparative analysis
1118 of surface-exposed virulence factors of Acinetobacter baumannii. BMC Genomics.
1119 2014;15:1020. Epub 2014/11/26. doi: 10.1186/1471-2164-15-1020. PubMed PMID:
1120 25422040; PubMed Central PMCID: PMCPMC4256060.

1121 39. Antunes LC, Visca P, Towner KJ. Acinetobacter baumannii: evolution of a global
1122 pathogen. Pathog Dis. 2014;71(3):292-301. Epub 2014/01/01. doi: 10.1111/2049-
1123 632X.12125. PubMed PMID: 24376225.
1124 40. Peleg AY, de Breij A, Adams MD, Cerqueira GM, Mocali S, Galardini M, et al. The
1125 success of acinetobacter species; genetic, metabolic and virulence attributes. PLoS ONE.
1126 2012;7(10):e46984. Epub 2012/11/13. doi: 10.1371/journal.pone.0046984. PubMed PMID:
1127 23144699; PubMed Central PMCID: PMC3483291.
1128 41. Meumann EM, Anstey NM, Currie BJ, Piera KA, Kenyon JJ, Hall RM, et al. Genomic
1129 epidemiology of severe community-onset Acinetobacter baumannii infection. Microb
1130 Genom. 2019;5(3). Epub 2019/02/27. doi: 10.1099/mgen.0.000258. PubMed PMID:
1131 30806611; PubMed Central PMCID: PMCPMC6487312.
1132 42. O'Leary NA, Wright MW, Brister JR, Ciufo S, Haddad D, McVeigh R, et al. Reference
1133 sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional
1134 annotation. Nucleic Acids Res. 2016;44(D1):D733-45. Epub 2015/11/11. doi:
1135 10.1093/nar/gkv1189. PubMed PMID: 26553804; PubMed Central PMCID:
1136 PMCPMC4702849.
1137 43. Comai S, Bertazzo A, Brughera M, Crotti S. Chapter Five - Tryptophan in health and
1138 disease. In: Makowski GS, editor. Advances in Clinical Chemistry. 95: Elsevier; 2020. p. 165-
1139 218.
1140 44. Costantini C, Bellet MM, Renga G, Stincardini C, Borghi M, Pariano M, et al.
1141 Tryptophan Co-Metabolism at the Host-Pathogen Interface. Front Immunol. 2020;11:67.
1142 Epub 2020/02/23. doi: 10.3389/fimmu.2020.00067. PubMed PMID: 32082324; PubMed
1143 Central PMCID: PMCPMC7001157.
1144 45. Altenhoff AM, Levy J, Zarowiecki M, Tomiczek B, Warwick Vesztrocy A, Dalquen DA,
1145 et al. OMA standalone: orthology inference among public and custom genomes and
1146 transcriptomes. Genome Res. 2019;29(7):1152-63. Epub 2019/06/27. doi:
1147 10.1101/gr.243212.118. PubMed PMID: 31235654; PubMed Central PMCID:
1148 PMCPMC6633268.
1149 46. Galperin MY, Kristensen DM, Makarova KS, Wolf YI, Koonin EV. Microbial genome
1150 analysis: the COG approach. Brief Bioinform. 2019;20(4):1063-70. Epub 2017/10/03. doi:
1151 10.1093/bib/bbx117. PubMed PMID: 28968633; PubMed Central PMCID: PMCPMC6781585.
1152 47. Kanehisa M, Sato Y, Kawashima M, Furumichi M, Tanabe M. KEGG as a reference
1153 resource for gene and protein annotation. Nucleic Acids Res. 2016;44(D1):D457-62. doi:
1154 10.1093/nar/gkv1070. PubMed PMID: 26476454; PubMed Central PMCID:
1155 PMCPMC4702792.
1156 48. Chen L, Yang J, Yu J, Yao Z, Sun L, Shen Y, et al. VFDB: a reference database for
1157 bacterial virulence factors. Nucleic Acids Res. 2005;33(Database issue):D325-8. Epub
1158 2004/12/21. doi: 10.1093/nar/gki008. PubMed PMID: 15608208; PubMed Central PMCID:
1159 PMCPMC539962.
1160 49. Snitkin ES, Zelazny AM, Montero CI, Stock F, Mijares L, Murray PR, et al. Genome-
1161 wide recombination drives diversification of epidemic strains of Acinetobacter baumannii.
1162 Proc Natl Acad Sci U S A. 2011;108(33):13758-63. doi: 10.1073/pnas.1104404108.
1163 50. Mirarab S, Warnow T. ASTRAL-II: coalescent-based species tree estimation with many
1164 hundreds of taxa and thousands of genes. Bioinformatics. 2015;31(12):i44-52. Epub
1165 2015/06/15. doi: 10.1093/bioinformatics/btv234. PubMed PMID: 26072508; PubMed
1166 Central PMCID: PMCPMC4765870.
1167 51. Mateo-Estrada V, Grana-Miraglia L, Lopez-Leal G, Castillo-Ramirez S. Phylogenomics
1168 Reveals Clear Cases of Misclassification and Genus-Wide Phylogenetic Markers for

1169 Acinetobacter. Genome Biol Evol. 2019;11(9):2531-41. Epub 2019/08/14. doi:
1170 10.1093/gbe/evz178. PubMed PMID: 31406982; PubMed Central PMCID: PMCPMC6740150.
1171 52. Qin J, Feng Y, Lu X, Zong Z. Precise Species Identification for Acinetobacter: a
1172 Genome-Based Study with Description of Two Novel Acinetobacter Species. mSystems.
1173 2021;6(3):e0023721. Epub 2021/06/02. doi: 10.1128/mSystems.00237-21. PubMed PMID:
1174 34061620; PubMed Central PMCID: PMCPMC8269215.
1175 53. Kim PS, Shin NR, Kim JY, Yun JH, Hyun DW, Bae JW. Acinetobacter apis sp. nov.,
1176 isolated from the intestinal tract of a honey bee, Apis mellifera. J Microbiol. 2014;52(8):639-
1177 45. Epub 2014/08/08. doi: 10.1007/s12275-014-4078-0. PubMed PMID: 25098562.
1178 54. Ellegaard KM, Engel P. Genomic diversity landscape of the honey bee gut microbiota.
1179 Nature communications. 2019;10(1):446. Epub 2019/01/27. doi: 10.1038/s41467-019-
1180 08303-0. PubMed PMID: 30683856; PubMed Central PMCID: PMCPMC6347622.
1181 55. McFrederick QS, Rehan SM. Characterization of pollen and bacterial community
1182 composition in brood provisions of a small carpenter bee. Molecular ecology.
1183 2016;25(10):2302-11. Epub 2016/03/08. doi: 10.1111/mec.13608. PubMed PMID: 26945527.
1184 56. Karah N, Haldorsen B, Hegstad K, Simonsen GS, Sundsfjord A, Samuelsen O, et al.
1185 Species identification and molecular characterization of Acinetobacter spp. blood culture
1186 isolates from Norway. The Journal of antimicrobial chemotherapy. 2011;66(4):738-44. Epub
1187 2011/03/12. doi: 10.1093/jac/dkq521. PubMed PMID: 21393175.
1188 57. Turton JF, Shah J, Ozongwu C, Pike R. Incidence of Acinetobacter species other than
1189 A. baumannii among clinical isolates of Acinetobacter: evidence for emerging species. J Clin
1190 Microbiol. 2010;48(4):1445-9. Epub 2010/02/26. doi: 10.1128/JCM.02467-09. PubMed
1191 PMID: 20181894; PubMed Central PMCID: PMCPMC2849580.
1192 58. Niu C, Clemmer KM, Bonomo RA, Rather PN. Isolation and characterization of an
1193 autoinducer synthase from Acinetobacter baumannii. J Bacteriol. 2008;190(9):3386-92. Epub
1194 2008/02/19. doi: 10.1128/JB.01929-07. PubMed PMID: 18281398; PubMed Central PMCID:
1195 PMCPMC2347373.
1196 59. Bhargava N, Sharma P, Capalash N. Quorum sensing in Acinetobacter: an emerging
1197 pathogen. Crit Rev Microbiol. 2010;36(4):349-60. Epub 2010/09/18. doi:
1198 10.3109/1040841X.2010.512269. PubMed PMID: 20846031.
1199 60. Clemmer KM, Bonomo RA, Rather PN. Genetic analysis of surface motility in
1200 Acinetobacter baumannii. Microbiology. 2011;157(Pt 9):2534-44. Epub 2011/06/28. doi:
1201 10.1099/mic.0.049791-0. PubMed PMID: 21700662; PubMed Central PMCID:
1202 PMCPMC3352170.
1203 61. Choudhary KS, Hudaiberdiev S, Gelencser Z, Goncalves Coutinho B, Venturi V, Pongor
1204 S. The organization of the quorum sensing luxI/R family genes in Burkholderia. Int J Mol Sci.
1205 2013;14(7):13727-47. Epub 2013/07/04. doi: 10.3390/ijms140713727. PubMed PMID:
1206 23820583; PubMed Central PMCID: PMCPMC3742214.
1207 62. Le Guillouzer S, Groleau MC, Deziel E. Two rsaM Homologues Encode Central
1208 Regulatory Elements Modulating Quorum Sensing in Burkholderia thailandensis. J Bacteriol.
1209 2018;200(14). Epub 2018/03/07. doi: 10.1128/JB.00727-17. PubMed PMID: 29507087;
1210 PubMed Central PMCID: PMCPMC6018367.
1211 63. Lopez-Martin M, Dubern JF, Alexander MR, Williams P. AbaM Regulates Quorum
1212 Sensing, Biofilm Formation, and Virulence in Acinetobacter baumannii. J Bacteriol.
1213 2021;203(8). Epub 2021/01/27. doi: 10.1128/JB.00635-20. PubMed PMID: 33495249;
1214 PubMed Central PMCID: PMCPMC8088503.
1215 64. Rumbo-Feal S, Gomez MJ, Gayoso C, Alvarez-Fraga L, Cabral MP, Aransay AM, et al.
1216 Whole transcriptome analysis of Acinetobacter baumannii assessed by RNA-sequencing

1217 reveals different mRNA expression profiles in biofilm compared to planktonic cells. PLoS
1218 One. 2013;8(8):e72968. Epub 2013/09/12. doi: 10.1371/journal.pone.0072968. PubMed
1219 PMID: 24023660; PubMed Central PMCID: PMCPMC3758355.
1220 65. Rumbo-Feal S, Perez A, Ramelot TA, Alvarez-Fraga L, Vallejo JA, Beceiro A, et al.
1221 Contribution of the A. baumannii A1S_0114 Gene to the Interaction with Eukaryotic Cells
1222 and Virulence. Front Cell Infect Microbiol. 2017;7:108. Epub 2017/04/20. doi:
1223 10.3389/fcimb.2017.00108. PubMed PMID: 28421168; PubMed Central PMCID:
1224 PMCPMC5376624.
1225 66. Puri AW, Schaefer AL, Fu Y, Beck DAC, Greenberg EP, Lidstrom ME. Quorum Sensing
1226 in a Methane-Oxidizing Bacterium. J Bacteriol. 2017;199(5). Epub 2016/12/21. doi:
1227 10.1128/JB.00773-16. PubMed PMID: 27994019; PubMed Central PMCID:
1228 PMCPMC5309911.
1229 67. Gonzalez O, Ortiz-Castro R, Diaz-Perez C, Diaz-Perez AL, Magana-Duenas V, Lopez-
1230 Bucio J, et al. Non-ribosomal Peptide Synthases from Pseudomonas aeruginosa Play a Role in
1231 Cyclodipeptide Biosynthesis, Quorum-Sensing Regulation, and Root Development in a Plant
1232 Host. Microb Ecol. 2017;73(3):616-29. Epub 2016/12/03. doi: 10.1007/s00248-016-0896-4.
1233 PubMed PMID: 27900439.
1234 68. Tomaras AP, Dorsey CW, Edelmann RE, Actis LA. Attachment to and biofilm
1235 formation on abiotic surfaces by Acinetobacter baumannii: involvement of a novel
1236 chaperone-usher pili assembly system. Microbiology. 2003;149(Pt 12):3473-84. Epub
1237 2003/12/10. doi: 10.1099/mic.0.26541-0. PubMed PMID: 14663080.
1238 69. Pakharukova N, Tuittila M, Paavilainen S, Malmi H, Parilova O, Teneberg S, et al.
1239 Structural basis for Acinetobacter baumannii biofilm formation. Proc Natl Acad Sci U S A.
1240 2018;115(21):5558-63. Epub 2018/05/08. doi: 10.1073/pnas.1800961115. PubMed PMID:
1241 29735695; PubMed Central PMCID: PMCPMC6003481.
1242 70. Colclough AL, Scadden J, Blair JMA. TetR-family transcription factors in Gram-
1243 negative bacteria: conservation, variation and implications for efflux-mediated antimicrobial
1244 resistance. BMC Genomics. 2019;20(1):731. Epub 2019/10/14. doi: 10.1186/s12864-019-
1245 6075-5. PubMed PMID: 31606035; PubMed Central PMCID: PMCPMC6790063.
1246 71. Cuthbertson L, Nodwell JR. The TetR family of regulators. Microbiol Mol Biol Rev.
1247 2013;77(3):440-75. Epub 2013/09/06. doi: 10.1128/MMBR.00018-13. PubMed PMID:
1248 24006471; PubMed Central PMCID: PMCPMC3811609.
1249 72. Tomaras AP, Flagler MJ, Dorsey CW, Gaddy JA, Actis LA. Characterization of a two-
1250 component regulatory system from Acinetobacter baumannii that controls biofilm formation
1251 and cellular morphology. Microbiology. 2008;154(Pt 11):3398-409. Epub 2008/10/30. doi:
1252 10.1099/mic.0.2008/019471-0. PubMed PMID: 18957593.
1253 73. Cerqueira GM, Kostoulias X, Khoo C, Aibinu I, Qu Y, Traven A, et al. A global virulence
1254 regulator in Acinetobacter baumannii and its control of the phenylacetic acid catabolic
1255 pathway. J Infect Dis. 2014;210(1):46-55. Epub 2014/01/17. doi: 10.1093/infdis/jiu024.
1256 PubMed PMID: 24431277.
1257 74. Sun D, Crowell SA, Harding CM, De Silva PM, Harrison A, Fernando DM, et al. KatG
1258 and KatE confer Acinetobacter resistance to hydrogen peroxide but sensitize bacteria to
1259 killing by phagocytic respiratory burst. Life Sci. 2016;148:31-40. Epub 2016/02/11. doi:
1260 10.1016/j.lfs.2016.02.015. PubMed PMID: 26860891; PubMed Central PMCID:
1261 PMCPMC4792659.
1262 75. Slauch JM. How does the oxidative burst of macrophages kill bacteria? Still an open
1263 question. Mol Microbiol. 2011;80(3):580-3. Epub 2011/03/08. doi: 10.1111/j.1365-
1264 2958.2011.07612.x. PubMed PMID: 21375590; PubMed Central PMCID: PMCPMC3109634.

1265 76. Nguyen GT, Green ER, Mecsas J. Neutrophils to the ROScue: Mechanisms of NADPH
1266 Oxidase Activation and Bacterial Resistance. Front Cell Infect Microbiol. 2017;7:373. Epub
1267 2017/09/12. doi: 10.3389/fcimb.2017.00373. PubMed PMID: 28890882; PubMed Central
1268 PMCID: PMCPMC5574878.
1269 77. Mortensen BL, Skaar EP. The contribution of nutrient metal acquisition and
1270 metabolism to Acinetobacter baumannii survival within the host. Front Cell Infect Microbiol.
1271 2013;3:95. Epub 2014/01/01. doi: 10.3389/fcimb.2013.00095. PubMed PMID: 24377089;
1272 PubMed Central PMCID: PMCPMC3859900.
1273 78. Eijkelkamp BA, Hassan KA, Paulsen IT, Brown MH. Investigation of the human
1274 pathogen Acinetobacter baumannii under iron limiting conditions. BMC Genomics.
1275 2011;12:126. Epub 2011/02/24. doi: 10.1186/1471-2164-12-126. PubMed PMID: 21342532;
1276 PubMed Central PMCID: PMCPMC3055841.
1277 79. El-Zaatari M, Chang YM, Zhang M, Franz M, Shreiner A, McDermott AJ, et al.
1278 Tryptophan catabolism restricts IFN-gamma-expressing neutrophils and Clostridium difficile
1279 immunopathology. J Immunol. 2014;193(2):807-16. Epub 2014/06/18. doi:
1280 10.4049/jimmunol.1302913. PubMed PMID: 24935925; PubMed Central PMCID:
1281 PMCPMC4091639.
1282 80. Sheldon JR, Skaar EP. Acinetobacter baumannii can use multiple siderophores for iron
1283 acquisition, but only acinetobactin is required for virulence. PLoS Pathog.
1284 2020;16(10):e1008995. Epub 2020/10/20. doi: 10.1371/journal.ppat.1008995. PubMed
1285 PMID: 33075115; PubMed Central PMCID: PMCPMC7595644 following competing interests:
1286 Dr. Eric Skaar served as a consultant to Shionogi Inc.
1287 81. Nairn BL, Lonergan ZR, Wang J, Braymer JJ, Zhang Y, Calcutt MW, et al. The Response
1288 of Acinetobacter baumannii to Zinc Starvation. Cell Host Microbe. 2016;19(6):826-36. Epub
1289 2016/06/10. doi: 10.1016/j.chom.2016.05.007. PubMed PMID: 27281572; PubMed Central
1290 PMCID: PMCPMC4901392.
1291 82. Jakubovics NS, Jenkinson HF. Out of the iron age: new insights into the critical role of
1292 manganese homeostasis in bacteria. Microbiology (Reading). 2001;147(Pt 7):1709-18. Epub
1293 2001/06/29. doi: 10.1099/00221287-147-7-1709. PubMed PMID: 11429449.
1294 83. Anjem A, Varghese S, Imlay JA. Manganese import is a key element of the OxyR
1295 response to hydrogen peroxide in Escherichia coli. Mol Microbiol. 2009;72(4):844-58. Epub
1296 2009/04/30. doi: 10.1111/j.1365-2958.2009.06699.x. PubMed PMID: 19400769; PubMed
1297 Central PMCID: PMCPMC2776087.
1298 84. Juttukonda LJ, Chazin WJ, Skaar EP. Acinetobacter baumannii Coordinates Urea
1299 Metabolism with Metal Import To Resist Host-Mediated Metal Limitation. mBio. 2016;7(5).
1300 Epub 2016/09/30. doi: 10.1128/mBio.01475-16. PubMed PMID: 27677795; PubMed Central
1301 PMCID: PMCPMC5050338.
1302 85. Krizova L, Maixnerova M, Sedo O, Nemec A. Acinetobacter albensis sp. nov., isolated
1303 from natural soil and water ecosystems. Int J Syst Evol Microbiol. 2015;65(11):3905-12. Epub
1304 2015/08/08. doi: 10.1099/ijsem.0.000511. PubMed PMID: 26245775.
1305 86. Zhao J, Zhu Y, Han J, Lin YW, Aichem M, Wang J, et al. Genome-Scale Metabolic
1306 Modeling Reveals Metabolic Alterations of Multidrug-Resistant Acinetobacter baumannii in a
1307 Murine Bloodstream Infection Model. Microorganisms. 2020;8(11). Epub 2020/11/20. doi:
1308 10.3390/microorganisms8111793. PubMed PMID: 33207684; PubMed Central PMCID:
1309 PMCPMC7696501.
1310 87. Zhu Y, Zhao J, Maifiah MHM, Velkov T, Schreiber F, Li J. Metabolic Responses to
1311 Polymyxin Treatment in Acinetobacter baumannii ATCC 19606: Integrating Transcriptomics
1312 and Metabolomics with Genome-Scale Metabolic Modeling. mSystems. 2019;4(1). Epub

1313 2019/02/13. doi: 10.1128/mSystems.00157-18. PubMed PMID: 30746493; PubMed Central
1314 PMCID: PMCPMC6365644.
1315 88. Farrugia DN, Elbourne LD, Hassan KA, Eijkelkamp BA, Tetu SG, Brown MH, et al. The
1316 complete genome and phenome of a community-acquired Acinetobacter baumannii. PLoS
1317 One. 2013;8(3):e58628. Epub 2013/03/26. doi: 10.1371/journal.pone.0058628. PubMed
1318 PMID: 23527001; PubMed Central PMCID: PMCPMC3602452.
1319 89. Durot M, Le Fevre F, de Berardinis V, Kreimeyer A, Vallenet D, Combe C, et al.
1320 Iterative reconstruction of a global metabolic model of Acinetobacter baylyi ADP1 using
1321 high-throughput growth phenotype and gene essentiality data. BMC Syst Biol. 2008;2:85.
1322 Epub 2008/10/09. doi: 10.1186/1752-0509-2-85. PubMed PMID: 18840283; PubMed Central
1323 PMCID: PMCPMC2606687.
1324 90. Nemec A. Metabolic and physiological characteristics of Acinetobacter species 2021.
1325 Available from: https://apps.szu.cz/anemec/Phenotype.pdf.
1326 91. Matsushita K, Toyama H, Ameyama M, Adachi O, Dewanti A, Duine JA. Soluble and
1327 Membrane-bound Quinoprotein D-Glucose Dehydrogenases of the Acinetobacter
1328 calcoaceticus: The Binding Process of PQQ to the Apoenzymes. Bioscience, Biotechnology,
1329 and Biochemistry. 1995;59(8):1548-55. doi: 10.1271/bbb.59.1548.
1330 92. Shen YQ, Bonnot F, Imsand EM, RoseFigura JM, Sjolander K, Klinman JP. Distribution
1331 and properties of the genes encoding the biosynthesis of the bacterial cofactor,
1332 pyrroloquinoline quinone. Biochemistry. 2012;51(11):2265-75. Epub 2012/02/14. doi:
1333 10.1021/bi201763d. PubMed PMID: 22324760; PubMed Central PMCID: PMCPMC3334298.
1334 93. Riquelme SA, Ahn D, Prince A. Pseudomonas aeruginosa and Klebsiella pneumoniae
1335 Adaptation to Innate Immune Clearance Mechanisms in the Lung. J Innate Immun.
1336 2018;10(5-6):442-54. Epub 2018/04/05. doi: 10.1159/000487515. PubMed PMID: 29617698;
1337 PubMed Central PMCID: PMCPMC6785651.
1338 94. Nemec A. 2021.
1339 95. Blumenthal HJ, Lucuta VL, Blumenthal DC. Specific enzymatic assay for D-glucarate in
1340 human serum. Anal Biochem. 1990;185(2):286-93. Epub 1990/03/01. doi: 10.1016/0003-
1341 2697(90)90294-j. PubMed PMID: 2187374.
1342 96. Lamichhane-Khadka R, Benoit SL, Maier SE, Maier RJ. A link between gut community
1343 metabolism and pathogenesis: molecular hydrogen-stimulated glucarate catabolism aids
1344 Salmonella virulence. Open Biol. 2013;3(12):130146. Epub 2013/12/07. doi:
1345 10.1098/rsob.130146. PubMed PMID: 24307595; PubMed Central PMCID:
1346 PMCPMC3877842.
1347 97. Rosas-Lemus M, Minasov G, Shuvalova L, Wawrzak Z, Kiryukhina O, Mih N, et al.
1348 Structure of galactarate dehydratase, a new fold in an enolase involved in bacterial fitness
1349 after antibiotic treatment. Protein Sci. 2020;29(3):711-22. Epub 2019/12/08. doi:
1350 10.1002/pro.3796. PubMed PMID: 31811683; PubMed Central PMCID: PMCPMC7021002.
1351 98. Reuter SE, Evans AM. Carnitine and acylcarnitines: pharmacokinetic, pharmacological
1352 and clinical aspects. Clin Pharmacokinet. 2012;51(9):553-72. Epub 2012/07/19. doi:
1353 10.2165/11633940-000000000-00000
1354 10.1007/BF03261931. PubMed PMID: 22804748.
1355 99. Zhu Y, Jameson E, Crosatti M, Schafer H, Rajakumar K, Bugg TD, et al. Carnitine
1356 metabolism to trimethylamine by an unusual Rieske-type oxygenase from human
1357 microbiota. Proc Natl Acad Sci U S A. 2014;111(11):4268-73. Epub 2014/03/05. doi:
1358 10.1073/pnas.1316569111. PubMed PMID: 24591617; PubMed Central PMCID:
1359 PMCPMC3964110.

1360 100. Bouvet PJM, Grimont PAD. Taxonomy of the Genus Acinetobacter with the
1361 Recognition of Acinetobacter baumannii sp. nov., Acinetobacter haemolyticus sp. nov.,
1362 Acinetobacter johnsonii sp. nov., and Acinetobacter junii sp. nov. and Emended Descriptions
1363 of Acinetobacter calcoaceticus and Acinetobacter lwoffii. International Journal of Systematic
1364 and Evolutionary Microbiology. 1986;36(2):228-40. doi: https://doi.org/10.1099/00207713-
1365 36-2-228.
1366 101. Breisch J, Waclawska I, Averhoff B. Identification and characterization of a carnitine
1367 transporter in Acinetobacter baumannii. Microbiologyopen. 2019;8(6):e00752. Epub
1368 2018/10/16. doi: 10.1002/mbo3.752. PubMed PMID: 30318737; PubMed Central PMCID:
1369 PMCPMC6562126.
1370 102. Wang CH, Zhao TX, Li M, Zhang C, Xing XH. Characterization of a novel Acinetobacter
1371 baumannii xanthine dehydrogenase expressed in Escherichia coli. Biotechnol Lett.
1372 2016;38(2):337-44. Epub 2015/11/07. doi: 10.1007/s10529-015-1986-y. PubMed PMID:
1373 26543035.
1374 103. Altenhoff AM, Glover NM, Train CM, Kaleb K, Warwick Vesztrocy A, Dylus D, et al. The
1375 OMA orthology database in 2018: retrieving evolutionary relationships among all domains of
1376 life through richer web and programmatic interfaces. Nucleic Acids Res. 2018;46(D1):D477-
1377 D85. Epub 2017/11/07. doi: 10.1093/nar/gkx1019. PubMed PMID: 29106550; PubMed
1378 Central PMCID: PMCPMC5753216.
1379 104. Knoten CA, Hudson LL, Coleman JP, Farrow JM, 3rd, Pesci EC. KynR, a Lrp/AsnC-type
1380 transcriptional regulator, directly controls the kynurenine pathway in Pseudomonas
1381 aeruginosa. J Bacteriol. 2011;193(23):6567-75. Epub 2011/10/04. doi: 10.1128/JB.05803-11.
1382 PubMed PMID: 21965577; PubMed Central PMCID: PMCPMC3232868.
1383 105. Bortolotti P, Hennart B, Thieffry C, Jausions G, Faure E, Grandjean T, et al. Tryptophan
1384 catabolism in Pseudomonas aeruginosa and potential for inter-kingdom relationship. BMC
1385 Microbiol. 2016;16(1):137. Epub 2016/07/09. doi: 10.1186/s12866-016-0756-x. PubMed
1386 PMID: 27392067; PubMed Central PMCID: PMCPMC4938989.
1387 106. Rohmer L, Hocquet D, Miller SI. Are pathogenic bacteria just looking for food?
1388 Metabolism and microbial pathogenesis. Trends Microbiol. 2011;19(7):341-8. Epub
1389 2011/05/24. doi: 10.1016/j.tim.2011.04.003. PubMed PMID: 21600774; PubMed Central
1390 PMCID: PMCPMC3130110.
1391 107. Fournier P-E, Vallenet D, Barbe V, Audic S, Ogata H, Poirel L, et al. Comparative
1392 Genomics of Multidrug Resistance in Acinetobacter baumannii. PLoS Genet. 2006;2(1):e7.
1393 doi: 10.1371/journal.pgen.0020007.
1394 108. Sato Y, Unno Y, Miyazaki C, Ubagai T, Ono Y. Multidrug-resistant Acinetobacter
1395 baumannii resists reactive oxygen species and survives in macrophages. Sci Rep.
1396 2019;9(1):17462. Epub 2019/11/27. doi: 10.1038/s41598-019-53846-3. PubMed PMID:
1397 31767923; PubMed Central PMCID: PMCPMC6877552.
1398 109. Chen YJ, Leung PM, Wood JL, Bay SK, Hugenholtz P, Kessler AJ, et al. Metabolic
1399 flexibility allows bacterial habitat generalists to become dominant in a frequently disturbed
1400 ecosystem. ISME J. 2021. Epub 2021/05/05. doi: 10.1038/s41396-021-00988-w. PubMed
1401 PMID: 33941890.
1402 110. Ren W, Rajendran R, Zhao Y, Tan B, Wu G, Bazer FW, et al. Amino Acids As Mediators
1403 of Metabolic Cross Talk between Host and Pathogen. Front Immunol. 2018;9:319. Epub
1404 2018/03/15. doi: 10.3389/fimmu.2018.00319. PubMed PMID: 29535717; PubMed Central
1405 PMCID: PMCPMC5835074.

1406 111. Chen Y, Guillemin GJ. Kynurenine pathway metabolites in humans: disease and
1407 healthy States. Int J Tryptophan Res. 2009;2:1-19. Epub 2009/01/01. doi: 10.4137/ijtr.s2097.
1408 PubMed PMID: 22084578; PubMed Central PMCID: PMCPMC3195227.
1409 112. Genestet C, Le Gouellec A, Chaker H, Polack B, Guery B, Toussaint B, et al. Scavenging
1410 of reactive oxygen species by tryptophan metabolites helps Pseudomonas aeruginosa
1411 escape neutrophil killing. Free Radic Biol Med. 2014;73:400-10. Epub 2014/06/15. doi:
1412 10.1016/j.freeradbiomed.2014.06.003. PubMed PMID: 24929180.
1413 113. Page AJ, Taylor B, Keane JA. Multilocus sequence typing by blast from de novo
1414 assemblies against PubMLST. Journal of Open Source Software. 2016;1(8):118. doi:
1415 10.21105/joss.00118.
1416 114. Bartual SG, Seifert H, Hippler C, Luzon MA, Wisplinghoff H, Rodriguez-Valera F.
1417 Development of a multilocus sequence typing scheme for characterization of clinical isolates
1418 of Acinetobacter baumannii. J Clin Microbiol. 2005;43(9):4382-90. Epub 2005/09/08. doi:
1419 10.1128/JCM.43.9.4382-4390.2005. PubMed PMID: 16145081; PubMed Central PMCID:
1420 PMCPMC1234098.
1421 115. Diancourt L, Passet V, Nemec A, Dijkshoorn L, Brisse S. The population structure of
1422 Acinetobacter baumannii: expanding multiresistant clones from an ancestral susceptible
1423 genetic pool. PLoS One. 2010;5(4):e10034. Epub 2010/04/13. doi:
1424 10.1371/journal.pone.0010034. PubMed PMID: 20383326; PubMed Central PMCID:
1425 PMCPMC2850921.
1426 116. Wang X, Qiao F, Yu R, Gao Y, Zong Z. Clonal diversity of Acinetobacter baumannii
1427 clinical isolates revealed by a snapshot study. BMC Microbiol. 2013;13:234. Epub
1428 2013/10/23. doi: 10.1186/1471-2180-13-234. PubMed PMID: 24144168; PubMed Central
1429 PMCID: PMCPMC3815251.
1430 117. Wilharm G, Skiebe E, Higgins PG, Poppel MT, Blaschke U, Leser S, et al. Relatedness of
1431 wildlife and livestock avian isolates of the nosocomial pathogen Acinetobacter baumannii to
1432 lineages spread in hospitals worldwide. Environ Microbiol. 2017;19(10):4349-64. Epub
1433 2017/09/20. doi: 10.1111/1462-2920.13931. PubMed PMID: 28925528.
1434 118. Cerezales M, Xanthopoulou K, Wille J, Bustamante Z, Seifert H, Gallego L, et al.
1435 Acinetobacter baumannii analysis by core genome multi-locus sequence typing in two
1436 hospitals in Bolivia: endemicity of international clone 7 isolates (CC25). Int J Antimicrob
1437 Agents. 2019;53(6):844-9. Epub 2019/04/01. doi: 10.1016/j.ijantimicag.2019.03.019.
1438 PubMed PMID: 30928684.
1439 119. Kumar S, Patil PP, Singhal L, Ray P, Patil PB, Gautam V. Molecular epidemiology of
1440 carbapenem-resistant Acinetobacter baumannii isolates reveals the emergence of blaOXA-
1441 23 and blaNDM-1 encoding international clones in India. Infect Genet Evol. 2019;75:103986.
1442 Epub 2019/07/31. doi: 10.1016/j.meegid.2019.103986. PubMed PMID: 31362071.
1443 120. Higgins PG, Hrenovic J, Seifert H, Dekic S. Characterization of Acinetobacter
1444 baumannii from water and sludge line of secondary wastewater treatment plant. Water Res.
1445 2018;140:261-7. Epub 2018/05/04. doi: 10.1016/j.watres.2018.04.057. PubMed PMID:
1446 29723815.
1447 121. Levy-Blitchtein S, Roca I, Plasencia-Rebata S, Vicente-Taboada W, Velasquez-Pomar J,
1448 Munoz L, et al. Emergence and spread of carbapenem-resistant Acinetobacter baumannii
1449 international clones II and III in Lima, Peru. Emerg Microbes Infect. 2018;7(1):119. Epub
1450 2018/07/05. doi: 10.1038/s41426-018-0127-9. PubMed PMID: 29970918; PubMed Central
1451 PMCID: PMCPMC6030224.
1452 122. Tomaschek F, Higgins PG, Stefanik D, Wisplinghoff H, Seifert H. Head-to-Head
1453 Comparison of Two Multi-Locus Sequence Typing (MLST) Schemes for Characterization of

1454 Acinetobacter baumannii Outbreak and Sporadic Isolates. PLoS One. 2016;11(4):e0153014.
1455 Epub 2016/04/14. doi: 10.1371/journal.pone.0153014. PubMed PMID: 27071077; PubMed
1456 Central PMCID: PMCPMC4829225.
1457 123. Higgins PG, Prior K, Harmsen D, Seifert H. Development and evaluation of a core
1458 genome multilocus typing scheme for whole-genome sequence-based typing of
1459 Acinetobacter baumannii. PLoS One. 2017;12(6):e0179228. Epub 2017/06/09. doi:
1460 10.1371/journal.pone.0179228. PubMed PMID: 28594944; PubMed Central PMCID:
1461 PMCPMC5464626.
1462 124. Gaiarsa S, Batisti Biffignandi G, Esposito EP, Castelli M, Jolley KA, Brisse S, et al.
1463 Comparative Analysis of the Two Acinetobacter baumannii Multilocus Sequence Typing
1464 (MLST) Schemes. Frontiers in microbiology. 2019;10:930. Epub 2019/05/28. doi:
1465 10.3389/fmicb.2019.00930. PubMed PMID: 31130931; PubMed Central PMCID:
1466 PMCPMC6510311.
1467 125. Chernomor O, Minh BQ, Forest F, Klaere S, Ingram T, Henzinger M, et al. Split
1468 diversity in constrained conservation prioritization using integer linear programming.
1469 Methods Ecol Evol. 2015;6(1):83-91. Epub 2015/04/22. doi: 10.1111/2041-210X.12299.
1470 PubMed PMID: 25893087; PubMed Central PMCID: PMCPMC4392707.
1471 126. Kurtz S, Phillippy A, Delcher AL, Smoot M, Shumway M, Antonescu C, et al. Versatile
1472 and open software for comparing large genomes. Genome Biol. 2004;5(2):R12. Epub
1473 2004/02/05. doi: 10.1186/gb-2004-5-2-r12. PubMed PMID: 14759262; PubMed Central
1474 PMCID: PMCPMC395750.
1475 127. Pritchard L, Glover RH, Humphris S, Elphinstone JG, Toth IK. Genomics and taxonomy
1476 in diagnostics for food security: soft-rotting enterobacterial plant pathogens. Analytical
1477 Methods. 2016;8(1):12-24. doi: 10.1039/C5AY02550H.
1478 128. Birikmen M, Bohnsack KE, Tran V, Somayaji S, Bohnsack MT, Ebersberger I. Tracing
1479 Eukaryotic Ribosome Biogenesis Factors Into the Archaeal Domain Sheds Light on the
1480 Evolution of Functional Complexity. Frontiers in microbiology. 2021;12:2598.
1481 129. Tettelin H, Masignani V, Cieslewicz MJ, Donati C, Medini D, Ward NL, et al. Genome
1482 analysis of multiple pathogenic isolates of Streptococcus agalactiae: implications for the
1483 microbial "pan-genome". Proc Natl Acad Sci U S A. 2005;102(39):13950-5. Epub 2005/09/21.
1484 doi: 10.1073/pnas.0506758102. PubMed PMID: 16172379; PubMed Central PMCID:
1485 PMCPMC1216834.
1486 130. Katoh K, Toh H. Recent Developments in the MAFFT Multiple Sequence Alignment
1487 Program. Brief Bioinform. 2008;9(4):286-98.
1488 131. Suyama M, Torrents D, Bork P. PAL2NAL: robust conversion of protein sequence
1489 alignments into the corresponding codon alignments. Nucleic Acids Res. 2006;34(Web
1490 Server issue):W609-12. Epub 2006/07/18. doi: 10.1093/nar/gkl315. PubMed PMID:
1491 16845082; PubMed Central PMCID: PMCPMC1538804.
1492 132. Nguyen LT, Schmidt HA, von Haeseler A, Minh BQ. IQ-TREE: a fast and effective
1493 stochastic algorithm for estimating maximum-likelihood phylogenies. Mol Biol Evol.
1494 2015;32(1):268-74. doi: 10.1093/molbev/msu300. PubMed PMID: 25371430; PubMed
1495 Central PMCID: PMCPMC4271533.
1496 133. Huson DH. SplitsTree: analyzing and visualizing evolutionary data. Bioinformatics.
1497 1998;14(1):68-73. PubMed PMID: 9520503.
1498 134. Altenhoff AM, Gil M, Gonnet GH, Dessimoz C. Inferring hierarchical orthologous
1499 groups from orthologous gene pairs. PLoS One. 2013;8(1):e53786. Epub 2013/01/24. doi:
1500 10.1371/journal.pone.0053786. PubMed PMID: 23342000; PubMed Central PMCID:
1501 PMCPMC3544860.

135. The Gene Ontology C. The Gene Ontology Resource: 20 years and still GOing strong. Nucleic Acids Res. 2019;47(D1):D330-D8. Epub 2018/11/06. doi: 10.1093/nar/gky1055. PubMed PMID: 30395331; PubMed Central PMCID: PMCPMC6323945.
136. Supek F, Bošnjak M, Škunca N, Šmuc T. REVIGO Summarizes and Visualizes Long Lists of Gene Ontology Terms. PLOS ONE. 2011;6(7):e21800. doi: 10.1371/journal.pone.0021800.
137. Yu NY, Wagner JR, Laird MR, Melli G, Rey S, Lo R, et al. PSORTb 3.0: improved protein subcellular localization prediction with refined localization subcategories and predictive capabilities for all prokaryotes. Bioinformatics. 2010;26(13):1608-15. Epub 2010/05/18. doi: 10.1093/bioinformatics/btq249. PubMed PMID: 20472543; PubMed Central PMCID: PMCPMC2887053.
138. King BR, Vural S, Pandey S, Barteau A, Guda C. ngLOC: software and web server for predicting protein subcellular localization in prokaryotes and eukaryotes. BMC research notes. 2012;5:351. Epub 2012/07/12. doi: 10.1186/1756-0500-5-351. PubMed PMID: 22780965; PubMed Central PMCID: PMCPMC3532370.
139. Shen HB, Chou KC. Gneg-mPLoc: a top-down strategy to enhance the quality of predicting subcellular localization of Gram-negative bacterial proteins. J Theor Biol. 2010;264(2):326-33. Epub 2010/01/23. doi: 10.1016/j.jtbi.2010.01.018. PubMed PMID: 20093124.
140. Kanehisa M, Sato Y, Morishima K. BlastKOALA and GhostKOALA: KEGG Tools for Functional Characterization of Genome and Metagenome Sequences. J Mol Biol. 2016;428(4):726-31. doi: 10.1016/j.jmb.2015.11.006. PubMed PMID: 26585406.
141. Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. 1997;25(17):3389-402.
142. Fu L, Niu B, Zhu Z, Wu S, Li W. CD-HIT: accelerated for clustering the next-generation sequencing data. Bioinformatics. 2012;28(23):3150-2. Epub 2012/10/13. doi: 10.1093/bioinformatics/bts565. PubMed PMID: 23060610; PubMed Central PMCID: PMCPMC3516142.
143. Gower JC, Warrens MJ. Similarity, Dissimilarity, and Distance, Measures of. Wiley StatsRef: Statistics Reference Online. Hoboken, New Jersey, US: John Wiley & Sons, Ltd.; 2014.
144. Tucker AT, Nowicki EM, Boll JM, Knauf GA, Burdis NC, Trent MS, et al. Defining gene-phenotype relationships in Acinetobacter baumannii through one-step chromosomal gene inactivation. mBio. 2014;5(4):e01313-14. Epub 2014/08/07. doi: 10.1128/mBio.01313-14. PubMed PMID: 25096877; PubMed Central PMCID: PMCPMC4128354.
145. Datsenko KA, Wanner BL. One-step inactivation of chromosomal genes in Escherichia coli K-12 using PCR products. Proc Natl Acad Sci U S A. 2000;97(12):6640-5. Epub 2000/06/01. doi: 10.1073/pnas.120163297. PubMed PMID: 10829079; PubMed Central PMCID: PMCPMC18686.
146. Tschech A, Pfennig N. Growth yield increase linked to caffeate reduction in Acetobacterium woodii. Archives of Microbiology. 1984;137(2):163-7. doi: 10.1007/BF00414460.

1546 **Supplementary Information**

1547

1548 **S1 Text. Supplementary Text**

1549 Contains supplemental sections covering additional information on the taxon set

1550 compilation, a statistical exploration of all protein-coding genes, genomes and orthologs in

1551 Set-R, details on the method and workflow of the ESGC prediction, and provides additional

1552 results and discussions for the predicted $ESGCs_{ACB}$ including several clusters not discussed in

1553 the main manuscript.

1554 (.docx)

1555 **S1 Fig. In Set-R there is a significant difference in the number of coding sequences between**

1556 **the ACB clade members and non-members.**

1557 (A) Comparison of the number of coding sequences (CDS) per genome between members

1558 and non-members of ACB clade across *Acinetobacter*. It reveals a significant difference. ACB

1559 clade members, on average, contain 14% more protein coding genes. (B) Correlation matrix

1560 for a range of summary statistics on genome level across SET-R. Colored cells indicate value

1561 of spearman correlation coefficient [-1,1]. The descriptive statistics analyzed are explained in

1562 S2 Table. (C) Phylogenetic diversity of the orthologous groups (OGs) calculated from the sum

1563 of branch lengths of the subtree spanned by the taxa represented in an OG**.** This distribution

1564 contrasts the taxa belonging to the ACB clade vs. the total phylogenetic diversity. Data

1565 points are colored black if the corresponding OG belongs to the set of genus-wide core

1566 genes that were also used for phylogeny reconstruction. OGs represented closer to the

1567 upper left corner are especially interesting as they are approaching ubiquitous presence

1568 within the ACB clade but are rare in the rest of the genus (colored orange for illustrative

1569 purposes).

1570 (PDF)

1571 **S2 Fig. Majority-rule consensus phylogeny of 232 Acinetobacter strains represented in SET-**

1572 **R.**

1573 A high resolution image of the majority-rule consensus dendrogram of the Set-R taxa as

1574 shown in Fig 2A. Branches supported by only two out of three partition trees are indicated

1575 with dashed lines, branches supported by only one partition are not resolved. Leaf labels

1576 colored in green indicate changed species assignments. Such changes can either (i) correct

1577 (i.e. the original species assignment was at odds with the species assignment based on

1578 phylogenetic and ANI evidences), (ii) newly specify (i.e. the original species assignment was

1579 set to "unknown" (sp.)) or (iii) de-specify (the original species assignment could not be

1580 confirmed by phylogenetic evidences, and no alternative assignment was possible. The

1581 species label was set to "sp.") the species assignments as retrieved from NCBI RefSeq at the

1582 time of download.

1583 (PDF)

1584 **S3 Fig. Maximum likelihood (ML) tree for Set-F based on partition 1 of the core gene set**

1585 **(296 proteins).**

1586 Branch labels denote percent bootstrap support. The newick strings for the ML trees from all

1587 three partitions are given in S3 Data. A high-resolution figure of the consensus tree is

1588 provided in S9 Fig.

1589 (PDF)

1590 **S4 Fig. Hierarchically clustered ANIm heatmap across Set-R combined with phylogenetic**

1591 **information reveals unknown species diversity within the ACB clade.**

1592 The figure shows the color-graded average nucleotide identity (ANIm, all vs all) of the

1593 genomes across SET-R (both axes). Strains with high genomic identity generate clusters of

1594 high sequence identity (>95%: increasingly saturated red; 95%: white, <95%: saturating blue)
1595 along the diagonal. These clusters and cutoffs are typically used for bacterial species
1596 delineation. In the ACB clade, we observe a large and distinct cluster with 98-100% sequence
1597 identity (rounded) for the *A. baumannii* species (top left). Following the diagonal to the
1598 bottom right corner, the strains of *A. seifertii* (2, n=2) and *A. nosocomialis* (3, n=5) are
1599 clustered similarly. The percent identities decrease, for the following two clusters of
1600 genomes with mainly *A. calcoaceticus* species assignments (4 and 5). Specifically, only three
1601 pairs in cluster 4 and one pair in cluster 5 reach the species threshold ANI. Next to putatively
1602 misidentified *A. calcoaceticus*, cluster 5 features a strain of the tentative species *A.*
1603 *oleivorans* (strain CIP 110421) as well as the unassigned *A.* sp. WC-141 and *A.* sp. NIPH 817
1604 suggesting the existence of undescribed species in the ACB clade. Two further genomes,
1605 here located between cluster 5 and cluster 6 are putatively mislabeled strains *A. baumannii*
1606 strain 573719 and *A. pittii* ANC 4050. These strains again potentially represent undescribed
1607 species in the ACB clade. Clusters 6 and 7 comprise the strains from *A. lactucae* and from *A.*
1608 *pittii*, respectively. S12 Table contains the full matrix of pairwise ANIm in a tabulated format.
1609 (PNG)

1610 **S5 Fig. Genomic regions of all ESGC along the genome of ATCC 19606.**

1611 Graphical representations of the genomic regions for each ESGC with RD > 0 (see Methods)
1612 along the genome of ATCC 19606 with abbreviated abundance profiles and functional
1613 annotations.
1614 (PDF)

1615 **S6 Fig. Growth of *A. baumannii* ATCC 19606, *A. calcoaceticus* DSM 30006 and *A. baylyi* on**
1616 **different carbon source.**
1617 *A. baumannii* ATCC 19606 (△), *A. calcoaceticus* (□) and *A. baylyi* ADP1 (○) were grown in

1618 mineral medium with 20 mM carnitine (A), gluconate (B) or D-malate (C) as carbon source.
1619 Each value is the mean of +/- S. E.M. of at least three independent measurements.
1620 (PDF)

1621 **S7 Fig. Ortholog abundance profiles and functional characterizations of *gdhA* and *gdhB*.**
1622 (PDF)

1623 **S8 Fig. Numbers of HOG innovations and losses at each node of the majority-rule**
1624 **consensus tree of Set-R.**
1625 A consensus tree representation of Set-R with each inner node of the tree annotated with
1626 the number of HOGs associated to it as well as the number of lost HOGs in the subsumed
1627 clade according to the rules of Dollo (+) Parsimony (see Methods). All HOGs are provided in
1628 S13 Table. The nodes are labeled with an incremental id. On the lineage of *A. baumannii* we
1629 used the following replacements in the manuscript: NODE_1 = "ACB+BR", NODE_8 =
1630 "ACB+LW", NODE_9 = "ACB+BA", NODE_10 = "ACB+HA", NODE_11 = "ACB", NODE_12 =
1631 "BNS", NODE_13 = "B". Tip labels are represented as NCBI RefSeq Identifiers. Tip labels also
1632 show unique assembly accession and, if applicable either clone type assignment or
1633 (corrected) species assignment in form of the first four letters of the species names.
1634 (PDF)

1635 **S9 Fig. Majority-rule consensus tree (dendrogram) for Set-F summarizing the information**
1636 **in the maximum likelihood trees based on the three partitions of the core gene set**.
1637 (PDF)

1638 **S10 Fig. Treemaps (1-5) of enrichment biological processes as generated by REVIGO (*cf.* S5**
1639 **Table).**
1640 (PDF)

1641 **S1 Data. Phylogenetic profiles and microsynteny plots in html format for the curated top**

1642 **150 ESGCs$_{ACB}$ across Set-F as produced by the Vicinator tool.**

1643 Files can also be downloaded in zip-format via figshare:

1644 https://figshare.com/s/caa9a6680bc90e004365 (temporary private link).

1645 (ZIP)

1646 **S2 Data. Phylogenetic profiles and microsynteny plots in html format for ESGC$_{ACB}$-0452 and**

1647 **the kynurenine pathway cluster in *P. aeruginosa* across a sample of all *Proteobacteria*.**

1648 Files can also be downloaded in zip-format via figshare:

1649 https://figshare.com/s/caa9a6680bc90e004365 (temporary private link).

1650 (ZIP)

1651 **S3 Data. ML trees for all three partitions in newick format including support values.**

1652 Files can also be downloaded in zip-format via figshare:

1653 https://figshare.com/s/caa9a6680bc90e004365 (temporary private link).

1654 (ZIP)

1655 **S1 Table. Full list of genomes used in this study.**

1656 The set of assemblies represented in Set-F is provided in tabular format as extracted from

1657 NCBI RefSeq. Among other, the table lists RefSeq assembly and sample accessions,

1658 taxonomy ids (on strain and species level if available), species labels, and strain labels,

1659 assembly status, submission date, submission institute, and ftp link to resource.

1660 (XLSX)

1661 **S2 Table. Representative genomes selected with corrections and descriptive statistics.**

1662 The subset of assemblies in Set-R is provided in tabular format. The table lists RefSeq

1663 assembly and sample accessions, taxonomy ids, species labels, and corrected labels (see

1664 main text), strain names, clone type assignments (via MLST and literature) sampling site, and

1665 sampling year (where available) of the strains. The table further features the results of our
1666 genome sequence-specific quantitative and qualitative analyses. The descriptive statistics
1667 are described in the documents' sheet 'ColumnLegends'.
1668 (XLSX)

1669 **S3 Table. Orthology matrix as calculated from the all vs. all orthology search across Set-R**
1670 **(OMA orthologous groups)**

1671 (1) The column header contains the labels of the NCBI RefSeq assembly accession of each
1672 genome. Row indices refer to the orthologous group identifier. The matrix fields contain the
1673 NCBI RefSeq protein sequence accession if an ortholog was identified and left empty
1674 otherwise. (2) List of orthologous groups identifiers constituting the core genome by our
1675 definition. These were used for the tree reconstruction
1676 (XLSX)

1677 **S4 Table. Phylogeny-based taxonomic classifications of the genomes in Set-F into the**
1678 **defined clades**

1679 Consists of 15 sheets in total. Each sheet lists the NCBI RefSeq assembly accession and taxon
1680 label (species and strain label) of the genomes that were phylogenetically classified to
1681 belong to the following clades: B clade (*A. baumannii*), S clade (*A. seifertii*), NO clade (*A.*
1682 *nosocomialis*), BNS clade (three aforementioned), L clade (*A. lactucae*), PI clade (*A. pittii*), CA
1683 clade (*A. calcoaceticus*), CPL clade (three aforementioned), ACB clade, HA clade, BA clade,
1684 LW clade, BR clade, QI clade, and the outgroup. Genome assignments to *A. baumannii* which
1685 were at odds with its phylogenetic placement in our study are indicated by a blue line. In the
1686 same way, we highlighted assignments to *A. pittii* in light orange and *A. calcoaceticus* in dark
1687 orange.
1688 (XLSX)

1689 **S5 Table: Results of consecutive GO term enrichment analysis**

1690 (1) shows the significantly enriched GO terms ($p_{adjusted} < 0.001$) for proteins in those HOGs at
1691 a node where an ortholog was detected in at least 50% of the subsumed taxa. If genes of the
1692 same HOG are the source for multiple enriched terms, only the most specific term (highest
1693 depth) was kept for the final results. The raw results of the GO term enrichment analysis are
1694 presented in (2). The results of the GO term enrichment analyses considering all HOGs at a
1695 node are provided in (3). Detailed explanations of the column headers are placed in the
1696 sheet 'ColumnLegends'. The test results from (3) were extracted to generate input tables for
1697 REVIGO (sheets 4-8). The resulting treemaps are shown in S10 Fig 1-5.
1698 (XLSX)

1699 **S6 Table. ESGC identification and underlying data.**

1700 This file gives detailed information about the ESGC identification and provides the underlying
1701 data. (1) gives for each gene along the genome of *A. baumannii* ATCC 19606 the number of
1702 taxa per clade (and per international clone type) that harbor an ortholog. (2) gives, for each
1703 gene, the input vector used for the dissimilarity calculations, the individual thresholds (5th
1704 percentile) and, as an example, the calculated dissimilarity between each gene and the gene
1705 immediately upstream. The full matrix containing all pairwise dissimilarity calculations for
1706 the prediction of clusters is deposited on figshare [10.6084/m9.figshare.16910974,
1707 https://figshare.com/s/caa9a6680bc90e004365 (temporary private link)]. (3) lists the identified
1708 graph components along with abundance statistics (median of the proteins in a component)
1709 across the clades (both absolute and relative), retention difference between ACB vs. non-
1710 ACB (RD), and CCD scores. (4) lists detailed contextual information for the components
1711 including all functional annotations from various sources. Detailed explanations of the

1712 column headers for all tables are placed in the sheet 'ColumnLegends'.

1713 (XLSX)

1714 **S7 Table. Summary statistics of collected experimental evidence found in the literature for**

1715 **the assimilation of >100 potential carbon sources for *A. baumannii*.**

1716 (XLSX)

1717 **S8 Table. List of all non-ACB genomes across Set-R that harbor $ESGC_{ACB}$ 0016 including**

1718 **meta information regarding sampling and isolation.**

1719 (PDF)

1720 **S9 Table. Assembly accessions and species/strain labels for the *Proteobacteria* sample**

1721 **(n=1363).**

1722 (XLSX)

1723 **S10 Table. List of primers used for Ab04 mutant deletions Δkyn mutant strains**.

1724 (XLSX)

1725 **S11 Table. In silico MLST classifications of the genomes in Set-F**.

1726 (1) Results are in of Oxford-scheme and (2) of Pasteur-scheme typing. Sheets (3-10) list NCBI

1727 RefSeq assembly accession and strain labels for all genomes in the *A. baumannii* clade where

1728 the MLST classification allowed a mapping to one of the 8 international clone types.

1729 (XLSX)

1730 **S12 Table. Raw values of the pairwise ANIm comparisons shown in S4 Fig.**

1731 (XLSX)

1732 **S13 Table. Hierarchical orthology matrix as calculated from the all vs. all orthology search**

1733 **across Set-R along the phylogenetic tree (OMA hierarchical orthologous groups, HOGs).**

1734 The column header contains the labels of the NCBI RefSeq assembly accession of each
1735 genome. The second column (LCA) indicates the associated inner node of the HOG in the
1736 reconstructed phylogenetic tree. The values either refer to the node label on the lineage of
1737 *A. baumannii* or, if the node is not on lineage, an incremental node id. The corresponding
1738 nodes together with summary statistics are provided in S8 Fig. Row indices refer to the
1739 hierarchical orthologous group ID. The matrix fields contain the NCBI RefSeq protein
1740 sequence accession if an ortholog was identified otherwise they are left empty. (2) lists the
1741 identifiers of HOGs with prevalence in at least 231 genomes. These were used for the
1742 definition of core components during cluster identification
1743 (XLSX)

1744 **S14 Table. Results of BLAST search for homologues of the QS cluster against non-**
1745 ***Acinetobacter* genomes.**
1746 (XLSX).

1747