Comparative genomic analysis reveals a monophyletic cold adapted *Arthrobacter* cluster from polar and alpine regions

Liang Shen[1, 2], Yongqin Liu[1, 2*], Baiqing Xu[1, 2], Ninglian Wang[2, 3], Sten Anslan[4], Ping Ren[1], Fei Liu[5], Yuguang Zhou[6] and Qing Liu[6]

[1]Key Laboratory of Tibetan Environment Changes and Land Surface Processes, Institute of Tibetan Plateau Research, Chinese Academy of Sciences, Beijing, 100085, China

[2]CAS Center for Excellence in Tibetan Plateau Earth Sciences, Beijing, 100085, China

[3]College of Urban and Environmental Science, Northwest University, Xian, 710069, China

[4]Braunschweig University of Technology, Zoological Institute, Braunschweig, Germany

[5]Institute of Microbiology, Chinese Academy of Sciences, Beijing, 100101, China

[6]China General Microbiological Culture Collection Center (CGMCC), Institute of Microbiology, Chinese Academy of Sciences, Beijing 100101, China

*Author for correspondence:

Yongqin Liu, yqliu@itpcas.ac.cn, Institute of Tibetan Plateau Research, Chinese Academy of Sciences, Beijing, 100101, China

Abstract

Decrease in the frequency of arginine and increase in lysine are the trends that have been identified in the genomes of cold adapted bacteria. However, some cold adapted taxa show only limited or no detectable changes in the frequencies of amino acid composition. Here, we examined *Arthrobacter* spp. genomes from a wide range of environments on whether the genomic adaptations can be conclusively identified across genomes of taxa from polar and alpine regions. Phylogenetic analysis with a concatenated alignment of 119 orthologous proteins revealed a monophyletic clustering of seven polar and alpine isolated strains. Significant changes in amino acid composition related to cold adaptation were exclusive to seven of the twenty-nine strains from polar and alpine regions. Analysis of significant indicator genes and cold shock genes also revealed that clear differences could only be detected in the same seven strains. These unique characteristics may result from a vast exchange of genome content at the node leading to the monophyletic cold adapted *Arthrobacter* cluster predicted by the birth-and-death model. We then experimentally validated that strains with significant changes in amino acid composition have a better capacity to grow at low temperature than the mesophilic strains.

Importance

Acquisition of novel traits through horizontal gene transfer at the early divergence of the monophyletic cluster may accelerate their adaptation to low temperature. Our study reached a clear relationship between adaptation to cold and genomic features and would advanced in understanding the ambiguous results produced by the previous

44 studies on genomic adaption to cold temperature.

45 **Keywords:** Cold adaptation, Genomic featrues, *Arthrobacter*, Polar and alpine
46 regions

47

48 **Introduction**

49 Separated by large distances and climatic barriers, high Arctic, Antarctic, and high
50 alpine regions represent extreme cold environments, which have been successfully
51 colonized by cold-adapted microorganisms (1-3). In the polar and alpine regions,
52 cold-adapted microorganisms play a key role in biogeochemical transformations such
53 as carbon, nitrogen, and iron, which have both local and global impacts (4-6); thus, it
54 is important to understand the adaptation strategy of microorganisms in the extreme
55 cold biome.

56

57 Temperature is a strong selective force that shapes the structure and function of
58 microorganisms (7). Thriving in cold environments requires cold-adapted bacteria to
59 synthesize enzymes that perform effectively at low temperatures, one of the major
60 strategies is the modification to enzymes (8-10). Cold-adapted enzymes possess
61 generally a number of amino acid changes that impart higher degree of structural
62 flexibility (fewer salt bridges by reduce arginine and proline contents) and higher
63 specific activity (more local mobility by increase asparagine, methionine and glycine
64 contents) at low temperatures than their mesophilic counterparts (11). The increased
65 low temperature activity of enzymes via change in amino acid composition was

validated by analysis the structural, kinetic and microcalorimetric of cold adapted enzymes, e.g. aminopeptidase, β-lactamases, and dienelactone hydrolase (GaDlh) (12-14).

Comparative genomic analyses to examine differences in amino acid composition toward cold adaptation came through the study of two methanogenic Archaea, *Methanogenium frigidum* and *Methanococcoides burtonii* from Ace Lake, Antarctica (15). Proteins from these cold-adapted Archaea were characterized by a higher proportion of non-charged polar amino acids, such as glutamine and threonine, and a lower proportion of hydrophobic amino acids, particularly leucine (15). The amino acid shifts toward increased enzyme flexibility, which confers catalytic efficiency and contribute to cold adaptation can be identified in genomes of *Psychromonas ingrahamii* 37, *Exiguobacterium sibiricum* 255-15, *Psychrobacter arcticus* 273-4, *Shewanella* spp. and *Glaciecola* spp. (16-19). However, these changes were limited in *Colwellia psychrerythraea* 34H, *Planococcus halocryophilus* Or1, *Rhodococcus* sp. JG3, *Arthrobacter* spp., *Actinotalea* sp. KRMCY2, *Polaromonas* sp. Eur3 1.2.1, *Paenisporosarcina* sp. Eur1 9.01.10, *Methylobacterium* sp. AL-11 and *Kocuria* sp. KROCY2 (13, 20-24). Even, no changes in amino acid composition was identified in *Desulfotalea psychrophila*, *Psychroflexus torquis* and *Arcticibacter* spp. (25-27). Thus, these analyses, which attempt to find specific adaptations to cold environments by comparing the amino acid composition seem to have resulted in inconsistent outcomes (22, 28). Furthermore, Csp (cold shock protein) genes may vary widely

88 across psychrophiles' genomes, and are even absent in some strains, such as
89 *Rhodoferax ferrireducens* T118 and *Methanolobus psychrophilus* R15. This indicates
90 an ambiguous correlation between the number of Csp genes and cold tolerance (9, 29).
91 It remains unknown why these genomic shifts toward cold adaptations are not
92 common features that shared by strains isolated across the alpine and polar regions.
93
94 In the present study, we aimed to verify whether there are genomic features that could
95 be conclusively identified by investigating the genomic patterns of *Arthrobacter* spp.
96 that are isolated from permanently cold environments. These strains share close last
97 common ancestry, and therefore differences observed between cold derived genomes
98 and the reference genomes are more likely to be the result of cold adaptation (22, 30).
99 Furthermore, we also investigated whether changes in the amino acid composition
100 towards cold adaptation can promote growth at low temperatures.
101
102 **Materials and Methods**
103 Sixteen *Arthrobacter* strains were isolated from the Tibetan Plateau (TP) using R2A
104 medium at 4 °C, (Supplementary Table S1). Growth curves at various temperatures
105 (25 °C, 5 °C and −1 °C) was measured in R2A broth. For the −1 °C temperature
106 treatment, the temperature was maintained with ice–water mixtures and by controlling
107 the ambient temperatures at 0 °C. The R2A broth remained liquid at −1 °C (27). Other
108 temperature treatments were sustained using a constant-temperature incubator. To
109 monitor growth, absorbance was measured at 600 nm on a Microplate Reader (MD

SpectraMax M5). L. Stokes (31) have suggested a one week period of incubation at 0 °C in order to define how well microorganisms are adapted to cold. However, in the present study we performed 10 days of incubation at −1 °C. The growth curve tests performed on strains were deposited at the CGMCC (China General Microbiological Culture Collection Center) under the accession numbers *Arthrobacter* sp. N199823 = CGMCC1.16197, *Arthrobacter* sp. 4R501 = CGMCC1.16194, *Arthrobacter* sp. B1805 = CGMCC1.16193, *Arthrobacter* sp. 08Y14 = CGMCC1.16198 and *Arthrobacter* sp. 9E14 = CGMCC1.16188. Strains *Arthrobacter* sp. A3 (CGMCC1.8987), *A.* alpinus CGMCC1.8950, *A. globiformis* CGMCC1.1894 and *A. luteolus* CGMCC1.1218 were from CGMCC. The whole genome shotgun sequences were deposited at DDBJ/ENA/GenBank under the BioProject PRJNA421662.

The genomic DNA of the sixteen strains were extracted using TIANamp Bacteria DNA Kit (TIANGEN, Beijing) following to the manufacturer's instructions. The concentration of genomic DNA was assessed with a NanoDrop spectrophotometer (2000c, Thermo Scientific, USA) and had an OD 260/280 ratio of 1.8–2.0. The DNA was stored in TE buffer (pH 8.0) for genome sequencing. Sequencing was performed using Illumina Hiseq 2000. Reads were assembled using SPAdes v3.11.1 with default options (32). As the algorithm is sensitive to sequencing errors, low-quality reads were filtered prior to *de novo* assembly using Fastp with default options (33).

Reference genomes were downloaded from NCBI in March 2017. The completeness

of genomes was calculated using CheckM v1.0.7 with options lineage_wf, -t 16, -x fna (34). rRNA genes were called using RNAmmer (v.1.2) (35). Genomes with a completeness of less than 96% and lack of extractable full-length 16S rRNA reads were removed. The resulting set of 39 reference *Arthrobacter* genomes were used for further analysis along with the 16 genomes obtained in this study. Of the 55 genomes, 29 were from extreme cold polar and alpine regions, and 26 were from non-extreme-cold environments (Supplementary Table S1).

The air temperatures (2 m from surface) were downloaded from the European Centre for Medium-Range Weather Forecasts (ECMWF) ERA-Interim reanalysis Database (http://apps.ecmwf.int/datasets/data/interim-full-moda/levtype=sfc/) (36). Periods of 2008-2017 were selected from ERA-Interim with a spatial resolution 1.5°*1.5°. For the strains of *A. woluwensis* NBRC 107840, *A. woluwensis* DSM 10495 and *A. luteolus* NBRC 107841 isolated from human body we used 37 °C as the environment temperature.

All the genomes were annotated simultaneously in the present study with RAST (Rapid Annotation using Subsystem Technology) (37). Calculation of amino acid composition was carried out with the PERL script 'aminoacidUsage.pl' (38). One-way analysis of variance (ANOVA) was used to examine the differences in amino acids composition. Statistical significance was considered at $\alpha \leq 0.05$.

For gene family clustering, *Microbacterium* sp. (No.7) was chosen as the out-group. *Microbacterium* is one of the closest relatives to the *Arthrobacter* genus (39) and it is placed right at the lineage outside the *Arthrobacter*. In general, out-groups that closely related to the in-group species are better suited for phylogeny reconstruction than distantly related out-groups (40). Gene families were clustered using FastOrtho software (--pv_cutoff 1-e5 --pi_cutoff 50 --pmatch_cutoff 50) (http://enews.patricbrc.org/fastortho/). At the amino acid level using an E value of 10^{-5} and ≥ 50% global amino acid identity threshold, a total of 36, 699 orthologs were identified. Among these, 148 were universal to all the genomes sampled and 119 were single-copy orthologs. The number of single-copy identified in the present study is approximately consistent to the study of D. H. Parks et al. (41) which found 120 shared single-copy proteins in bacteria. The 119 mono-copy orthologs were then concatenated using custom-made PERL scripts. As a first step for a genome tree construction, the concatenated orthologous genes were aligned at the amino acid sequence level using Muscle software v3.8.31 (42). Non-conserved segments in the alignments were then trimmed using the Gblock v0.9b (43) to discard all gap-containing columns (-b1 = 50 -b4 = 5, other parameters were set as default). As a second step, probabilistic phylogenetic approaches were used to analyze the concatenation data (30, 366 sites) of the 119 orthologs. The PTHREADS version of RAxML v8.2.4 (-f a –x 12345 –p 12345 -#100 -m GTRGAMMAI) and IQ-TREE v1.6.0 (-b 1000 –m GTR+I+G4) were used to construct a Maximum Likelihood phylogenetic tree (44, 45). The MPI version of Mrbayes v3.2.6 (mcmc nchains = 16

176 burnin = 0.25, samplefreq = 100, Ngen = 10000000, lset nst = 6 rates = gamma) was
177 used to construct a Bayesian phylogenetic tree (46). As the evolutionary models for
178 different sites in multi-gene concatenated alignments may differ, PartitionFinder
179 software v2.1.0 was used to determine the best-fit partitioning scheme for RAxML
180 and Mrbayes (47) with default settings. The resultant trees were embellished with
181 Adobe Illustrator CS6 and iTOL v3 (48).
182
183 Ordinations and statistical analyses were performed using the vegan package v2.4.4
184 (49) using R v3.3.3. Genes significantly associated with cold and temperate
185 environments were calculated by Indicator Species Analysis as implemented in the R
186 library labdsv (http://ecology.msu.montana.edu/labdsv/R/). Significance was
187 calculated through random reassignment of groups with 1,000 permutations.
188
189 The 55 *Arthrobacter* genomes were used for ancestral reconstruction; the out-group
190 species *Microbacterium* sp. (No.7) was not included because reconstruction of
191 ancestral genome content using COUNT v9.1106 does not require out-group species
192 (50). The COUNT software uses birth-and-death models to identify the rates of gene
193 deletion, duplication, and loss in each branch and node of a phylogenetic tree. We
194 used the pan-genome matrices (Supplementary Table S2) and the phylogenetic
195 birth-and-death model implemented in COUNT, to reconstruct the ancestral genome
196 content of *Arthrobacter* species. Ancestral history reconstruction was performed by
197 posterior probabilities: one hundred rounds of rate optimization were computed with a

198 convergence threshold of 10^{-3} prior to ancestral reconstruction, other parameters were
199 set as default; Horizontally-transferred genes (HTgenes) were identified using a
200 threshold of probability of gain higher value than 0.95 at the destination node and
201 excluding gains occurring in the last common ancestor with a probability higher than
202 0.5, as suggested by Oliveira and colleagues (51).
203
204 **Results**
205 ***Distribution of Arthrobacter strains along their phylogenetic clade***
206 We constructed phylogenetic clustering based on concatenated alignment of 119
207 orthologous to yield a high-resolution tree. The 55 *Arthrobacter* strains were clustered
208 into three main lineages (lineage 1, 2 and 3 in Fig. 1a; Fig. S1). Strains isolated from
209 polar and alpine environments were mixed in lineage 1 and 2 with the reference
210 strains. For example, despite being isolated from different environments, strains
211 *Arthrobacter* sp. Y81 from Tibetan Plateau (TP) lake, *Arthrobacter* sp. TB 26 from
212 Antarctica marine sponge (52), *Arthrobacter* sp. FB24 from soil in Seymour, Indiana
213 (53) and *Arthrobacter* sp. SPG23 from contaminated soil at the Ford Motor Company
214 site in Genk, Belgium (54), clustered together in the lineage 1. Strains of *Arthrobacter*
215 sp. Soil782 and *Arthrobacter* sp. H5 were located together in lineage 2, despite the
216 former was isolated from plant material and the latter was isolated from Antarctic soil
217 (24, 55). However, strains in lineage 3, in the middle of the phylogenetic tree (Fig. 1a),
218 were all isolated from polar or alpine environments. These included strains of
219 *Arthrobacter* sp. GMC3, *Arthrobacter* sp. A3, *Arthrobacter* sp. N199823, and *A.*

alpinus ERGS4-06 (isolated from TP lake, permafrost, ice core and glacial stream water); *A. alpinus* R3-8 (isolated from Antarctic soil); *Arthrobacter* sp. PAMC 25486 (isolated from Arctic soil) and *A. alpinus* DSM22274 (isolated from alpine soil).

Based on the clustering results, we classified these 55 strains into three groups for comparative genomic analysis: group A comprised the 26 reference strains isolated from non-extreme-cold environments (e.g. rhizosphere soil, plant leaf surface and blood culture out of polar or alpine) in lineages 1 and 2; group B comprised 22 strains isolated from cold environments (polar or alpine) in lineages 1 and 2; group C comprised 7 strains isolated from cold environments in lineage 3 (Fig. 1a). We detected a decrease in the frequency of arginine and an increase in lysine, which occurred exclusively to the genomes of strains belonging to group C (Fig. 1b). The air temperatures (2 m from surface) of the polar and alpine strains habitats was significantly lower compared with the references strains habitat (ANOVA, $P < 0.005$, $F = 4.052$, Fig. 1c).

Strain growth at different temperatures

Strains isolated from Tibetan Plateau showed variable growth patterns, but in general grew better at low temperatures (5 °C, -1 °C, Fig. 2) than the references strains. Strains in group A and B grew faster than strains in group C in exponential phase before 36 h at 25 °C (Fig. 2a). At 5 °C, strains in group C tended to grow faster than strains in group A and group B in exponential phase before 144 h except strain

242 *Arthrobacter* sp. 4R501; strains in group B grew faster than strains in group A (Fig.
243 2b). At -1 °C, strains in group C tended to grow faster than all the strains in group A
244 (growth was inhibited) and group B in exponential phase before 240 h (mean OD600
245 of group A, B and C was 0.00, 0.02 and 0.25 at 240 h, respectively, Fig. 2c). The
246 result is in consistent with the study on snow-bacteria of the Tibetan Plateau which
247 revealed the adaption to cold environments was the result from the expansion of their
248 minimum growth-temperature (56).
249
250 ***Features in amino acid composition***
251 The pattern of amino acid distribution in *Arthrobacter* spp. displayed an overall
252 similar trends in their genomes, with alanine being most abundant, followed by
253 leucine, glycine, and valine, while methionine, tryptophan and to a lesser extent
254 cysteine were infrequent. However, compared against reference genomes, we found a
255 decrease in the frequency of arginine and an increase in lysine have occurred
256 exclusively in the genomes of strains belonging to group C. The shift in composition
257 of arginine and lysine is closely related with survival strategies of psychrophiles (11).
258 Then, we calculated the composition of the twenty common amino acids in whole
259 proteins to determine the differences in their frequencies between the three groups (A,
260 B and C). For group C, significant changes in twenty amino acids were apparent
261 (ANOVA, $P < 0.005$, $F = 4.1491$, Fig. 3a). Of the amino acids that increased in group
262 C, one is positively charged (lysine) and seven are uncharged (tryptophan, threonine,
263 serine, asparagine, methionine, isoleucine and phenylalanine) (Fig. 3a). Of the amino

acids that decreased in group C, three are charged (arginine, glutamic acid and aspartic acid), and three are uncharged (proline, glycine and alanine) (Fig. 3a). By exhibiting a number of amino acid changes that impart higher degree of structural flexibility (fewer salt bridges by reduce arginine, glutamic acid, aspartic acid and proline contents) and higher specific activity (more local mobility by increase asparagine, methionine and lysine contents) (11, 57), the enzyme activities of group C strains were predicted to be increased. The resulted increase in flexibility and decrease in thermodynamic stability were consistent with the experimental data that strains in group C grew better at low temperatures while weakly at a higher temperature compared to the reference strains. For group B, there were only changes in glycine (ANOVA, decrease with P = 0.0245, F = 4.052) and threonine (increase with P = 0.0366, F = 4.052) (Fig. 3b). The differences in arginine, lysine and proline composition were not significant in group B.

Differences in the distribution of gene families between Arthrobacter strains

We performed a multivariate assessment of gene composition classified at the level of function. Ordination of functional genes using two-dimensional nonmetric multidimensional scaling (NMDS) revealed a clear separation of group C, while group A and B were not clearly separated (Fig. 4a). The PERMANOVA analysis with 1000 permutations (58) showed significant difference between strains of group A and C at a functional level (Fig. 4a, P = 0.001, F = 6.492). To remove the potential difference introduced by the distance between lineage 1 and lineage 2, we performed

the PERMANOVA analysis between group A and B separately in lineage 1 and lineage 2. The results revealed no significant differences between cold-environment derived strains and the reference strains in lineage 1 and lineage 2 (P = 0.505 and P = 0.171, respectively).

We further used indicator gene analysis to statistically define the characteristic genes contributing to the differences between the three groups. When group A was compared with group C, a total of 176 and 304 significant functional indicators were found in group A and C, respectively (Fig. 4b, ANOVA, $P < 0.05$, supplementary Table S3). In contrast, when group A was compared with group B, only 25 and 40 significant functional indicators were found, respectively (Fig. 4b, supplementary Table S3). The group C indicator genes were mainly affiliated to the functional category of Carbohydrates, Amino Acids and Derivatives and Cofactors/Vitamins/Prosthetic Groups/Pigments (Fig. 4b, supplementary Table S4).

Cold shock genes of Arthrobacter strains

All of the genomes in group C contained two predicted cold shock genes (one copy of *cspA* and one *cspC*). All of the strains in group A and B had an extra 1 to 5 copies of *cspA*, except strain *Arthrobacter* sp. H5 (Supplementary Table S2). We generated alignments and a maximum likelihood tree for cold shock genes. The phylogenetic analyses showed that *cspA* and *cspC* genes of strains in group C were monophyletic, while cold shock proteins (Csps) from group B were polyphyletic, interleaving with

group A (Fig. 4c).

Ancestral reconstruction of Arthrobacter strains and the dynamics of genome content

We obtained a phylogenomic tree of the 55 *Arthrobacter* strains to reconstruct the ancestral genome content using three different phylogenomic approaches (IQ-TREE, RAxML and MrBayes; Fig. 5a and Fig. S1). The phylogenetic birth-and-death model imposed on the phylogenomic tree revealed a steady trend towards genome expansion since the most recent common ancestor indicated by N54 (~ 2,460 gene families, Fig. 5a). The extant *Arthrobacter* genomes (2,867 to 4,521 gene families, average ~3,500 gene families, Fig. 5a) exhibited a complicated evolutionary path to net genome expansion. Our attention was mainly focused on nodes of N27, N33 and N53 leading to lineage 1 and lineage 2 and lineage 3 (refer to group C), because strains in group C could be distinguished from group A with respect to overall amino acid composition, NMDS analysis and indicator genes. In contrast, these differences could not be detected between group A and B. Our results showed that node N33 that gained 876 genes, was more divergent than nodes N27 and N53 which gained 131 and 91 genes, respectively (Fig. 5b). The average genes gained per branch, showed the same trend (N33 = 125, N27 = 5 and N53 = 4), indicating that the significant difference between N33 and others was not due to the lineage size. The number of genes lost at node N33 was also higher compared to N27 and N53 (the number of lost genes: 362, 47 and 40,

330 respectively; Fig. 5b). Almost half (429 of 876) of these genes were identified as
331 HTgenes for node N33, while none of them were identified as HTgenes at N27, N33
332 and N53. Genes gained at node N33 were related to cofactors/vitamins/prosthetic
333 groups/pigments (5.6%), membrane transport (3.7%), carbohydrates and amino acids
334 (3.7%) and derivatives (3.7%), but the function of large proportion of gene remained
335 unknown (65.2%) (Fig. 5c).
336

337 **Discussion**

338 In the current study, we found that not all *Arthrobacter* strains isolated from polar and
339 alpine regions exhibited significant detectable changes in the genome composition.
340 Although changes in amino acid composition towards cold adaptation can be widely
341 identified in psychrophilic strains (9, 11), strains in our group B which were isolated
342 from cold environments and able to grow at 5 °C, exhibited no significance in
343 changes in amino acid composition. This has also been the case of Antarctic
344 *Arthrobacter* isolates, for example, strains *Arthrobacter* sp. FB24, *Arthrobacter* sp.
345 Br18 and *Arthrobacter* sp. H5 were not located in lineage 3 in Fig. 1a, in which no
346 remarkable genomic features were identified (24). Thus, in certain cases, amino acid
347 shifts are limited and even no changes in amino acid composition can be identified,
348 despite the strains were able to grow at low temperatures. However, the trends in
349 amino acid composition in group C exhibited adaptations to cold environments in
350 terms of genome-wide amino acid composition, which were consistent with those of
351 cold adapted bacteria (9, 11). Thus, conserved genomic traits were exclusive to a

certain group in *Arthrobacter*.

Strains in group C shared more genomic traits with the established rules for protein adaptation to cold than that of in group B (15, 17, 18, 57, 59). Thus, strains in group C may represent a "better adapted" genome type which were better adapted to their native cold habitats, this is supported by the faster growth of strains in group C at 5 and –1 °C than strains in group A and B. The wide abundance of group C indicator genes in the functional categories revealed that psychrophilic lifestyle is most likely conferred by a collection of synergistic changes in overall genome content rather than a unique set of genes (23). Grouping of bacteria isolated from cold environments based on their growth pattern at low temperature and phylogenetic clustering helped in the identification of conserved genomic traits of cold adaptation.

Cold shock proteins (Csp) regulate the cold shock response and play a critical role in bacterial growth at low temperatures (Jones and Inouye, 1994; Hébraud and Potier, 1999). Psychrophiles vary widely in the number of Csp genes present in their genomes, indicating a weak correlation between the number of Csp genes and cold tolerance (9). In the present study, we did not find any increase in the number of Csp genes in group C, which contains strains that grew well at -1 °C. On the contrary, strains in group C had fewer Csp genes than those in groups A and B despite the group C strains exhibiting faster growth rates atlow temperatures. Thus, increase in number of Csp genes may not be the strategy for cold adaptation of *Arthrobacter* spp.

374

375 Strains in group C shared conserved genomic traits to cold adaptation and the
376 clustering of these strains in one monophyletic lineage suggests that they all have
377 evolved in cold environments and possess similar strategies to remain active and
378 survive low/freezing temperatures. The dynamic historical pattern of *Arthrobacter*
379 genomes is concordant with the emerging view that genomes evolve through a
380 dynamic process of expansion and streamlining (60-62). However, the evolution of a
381 cluster of strains which can be clearly separated from their relatives has rarely been
382 studied. Our results showed that the N33 node (which leads to the strains in group C)
383 exhibited early vast genome dynamics, which may play an important role in
384 promoting the growth of these strains at low temperatures. The result is consistent
385 with the study of Allen et al. (2009), which revealed the genome plasticity of *M.*
386 *burtonii* that have enabled adaptations to cold environments.Also, an inter-order
387 horizontal gene transfer event enabled the catabolism of compatible solutes by
388 *Colwellia psychrerythraea* 34H, which provided a selective advantage in cold (63).
389 This dynamic genome pattern is also in agreement with the general pattern of virtual,
390 higher taxa in Archaea and *Streptococcus* genomes, which have the key mechanisms
391 to help related taxa inhabit new niches (60, 61, 64). Because generation time in cold
392 environments is generally longer, horizontal gene transfer may be more effective for
393 acquiring beneficial traits rapidly (26, 65-68). This is experimentally validated by our
394 results that species belonging to group C, which exclusivity exhibited vast dynamic in
395 genome content, have a better capacity to grow at low temperature than their cold

396 environment isolated counterparts.

397

398 Based on the result of reconstruction of most recent common ancestor, it is suggested
399 that group C strains may be cold evolutionary legacies. Cold adaptations are
400 superimposed on pre-existing microorganisms and the temperature-dependent
401 distribution of bacteria may not result from widespread contemporary dispersal but is
402 an ancient evolutionary legacy, as revealed by evolutional analysis of cold desert
403 cyanobacteria and thermal traits of *Streptomyces* sister-taxa (69, 70). Strain *A. alpinus*
404 DSM 22274 in group C is a new species isolated from Alps (71). Many other new
405 bacteria species have been described from polar and alpine regions further suggesting
406 that the level of cold origin taxa could be considerable, and these may represent
407 endemic species (72-79).

408

409 **Conclusions**

410 Changes in genome composition and obtaining new genes via horizontal gene transfer
411 may not essential for bacteria to survive in cold environments. However, for strains
412 belonging to group C in our study, their adaptation to cold is accelerated by the
413 acquisition of novel traits through horizontal gene transfer (51). Our results indicate
414 that growth at 5 °C may not require significant changes in genome content, but
415 genomic modification seems to be essential for *Arthrobacter* spp. to grow well at
416 subzero temperature (-1 °C in this study). We found that significantly conserved
417 genomic traits could be detectable across the cold adapted strains that growth quickly

418 at subzero temperature.

419

420 **Conflict of Interest**

421 The authors declare no conflict of interest.

422

423 **Acknowledgements**

424 This study was financially supported by the National Natural Science Foundation of
425 China (Grant Nos. 41701085 and 41425004). Dr. Wei Zhu from the Beijing Institute
426 of Genomics CAS is thanked for his bioinformatics assistance. Dr. Qilong Qin from
427 Shandong University is thanked for his valuable suggestions.

428

429 **References**

430 1. Piette F, Struvay C, Feller G. 2011. The protein folding challenge in psychrophiles: facts and
431 current issues. Environ Microbiol 13:1924–1933.
432 2. Boetius A, Anesio AM, Deming JW, Mikucki JA, Rapp JZ. 2015. Microbial ecology of the
433 cryosphere: sea ice and glacial habitats. Nat Rev Microbiol 13:677–690.
434 3. Jungblut AD, Lovejoy C, Vincent WF. 2010. Global distribution of cyanobacterial ecotypes in
435 the cold biosphere. ISME J 4:191-202.
436 4. Hodson A, Anesio AM, Tranter M, Fountain A, Osborn M, Priscu J, Laybourn-Parry J, Sattler
437 B. 2008. Glacial ecosystems. Ecol Monogr 78:41-67.
438 5. Anesio AM, Laybourn-Parry J. 2012. Glaciers and ice sheets as a biome. Trends in Ecology &
439 Evolution 27:219-225.
440 6. Segawa T, Ishii S, Ohte N, Akiyoshi A, Yamada A, Maruyama F, Li Z, Hongoh Y, Takeuchi N.
441 2014. The nitrogen cycle in cryoconites: naturally occurring nitrification-denitrification
442 granules on a glacier. Environ Microbiol 16:3250-62.
443 7. Bahram M, Hildebrand F, Forslund SK, Anderson JL, Soudzilovskaia NA, Bodegom PM,
444 Bengtsson-Palme J, Anslan S, Coelho LP, Harend H, Huerta-Cepas J, Medema MH, Maltz
445 MR, Mundra S, Olsson PA, Pent M, Polme S, Sunagawa S, Ryberg M, Tedersoo L, Bork P.
446 2018. Structure and function of the global topsoil microbiome. Nature 560:233-237.
447 8. Chattopadhyay MK, Reddy GS, Shivaji S. 2014. Psychrophilic Bacteria: Biodiversity,
448 Molecular Basis of Cold Adaptation and Biotechnological Implications. Curr Biotechno
449 3:100-116.

9. Siddiqui KS, Williams TJ, Wilkins D, Yau S, Allen MA, Brown MV, Lauro FM, Cavicchioli R. 2013. Psychrophiles. Annu Rev Earth Planet Sci 41:87-115.
10. Grzymski JJ, Carter BJ, DeLong EF, Feldman RA, Ghadiri A, Murray AE. 2006. Comparative genomics of DNA fragments from six Antarctic marine planktonic bacteria. Appl Environ Microbiol 72:1532-41.
11. De Maayer P, Anderson D, Cary C, Cowan DA. 2014. Some like it cold: understanding the survival strategies of psychrophiles. EMBO Rep 15:508-517.
12. Michaux C, Massant J, Kerff F, Frere JM, Docquier JD, Vandenberghe I, Samyn B, Pierrard A, Feller G, Charlier P, Van Beeumen J, Wouters J. 2008. Crystal structure of a cold-adapted class C beta-lactamase. FEBS J 275:1687-97.
13. Huston AL, Haeggstrom JZ, Feller G. 2008. Cold adaptation of enzymes: structural, kinetic and microcalorimetric characterizations of an aminopeptidase from the Arctic psychrophile *Colwellia psychrerythraea* and of human leukotriene A(4) hydrolase. Biochim Biophys Acta 1784:1865-72.
14. Hashim NHF, Mahadi NM, Illias RM, Feroz SR, Abu Bakar FD, Murad AMA. 2018. Biochemical and structural characterization of a novel cold-active esterase-like protein from the psychrophilic yeast Glaciozyma antarctica. Extremophiles 22:607-616.
15. Saunders NFW, Thomas T, Curmi PMG, Mattick JS, Kuczek E, Slade R, Davis J, Franzmann PD, Boone D, Rusterholtz K, Feldman R, Gates C, Bench S, Sowers K, Kadner K, Aerts A, Dehal P, Detter C, Glavina T, Lucas S, Richardson P, Larimer F, Hauser L, Land M, Cavicchioli R. 2003. Mechanisms of thermal adaptation revealed from the genomes of the Antarctic Archaea *Methanogenium frigidum* and *Methanococcoides burtonii*. Genome Res 13:1580-1588.
16. Rodrigues DF, Ivanova N, He Z, Huebner M, Zhou J, Tiedje JM. 2008. Architecture of thermal adaptation in an *Exiguobacterium sibiricum* strain isolated from 3 million year old permafrost: a genome and transcriptome approach. BMC Genomics 9:547.
17. Ayala-del-Río HL, Chain PS, Grzymski JJ, Ponder MA, Ivanova N, Bergholz PW, Di Bartolo G, Hauser L, Land M, Bakermans C, Rodrigues D, Klappenbach J, Zarka D, Larimer F, Richardson P, Murray A, Thomashow M, Tiedje JM. 2010. The genome sequence of *Psychrobacter arcticus* 273-4, a psychroactive Siberian permafrost bacterium, reveals mechanisms for adaptation to low-temperature growth. Appl Environ Microbiol 76:2304-12.
18. Zhao J, Deng Y, Manno D, Hawari J. 2010. *Shewanella* spp. genomic evolution for a cold marine lifestyle and in-situ explosive biodegradation. PloS One 5:e9109.
19. Qin QL, Xie BB, Yu Y, Shu YL, Rong JC, Zhang YJ, Zhao DL, Chen XL, Zhang XY, Chen B, Zhou BC, Zhang YZ. 2014. Comparative genomics of the marine bacterial genus *Glaciecola* reveals the high degree of genomic diversity and genomic characteristic for cold adaptation. Environ Microbiol 16:1642-1653.
20. Mykytczuk NCS, Foote SJ, Omelon CR, Southam G, Greer CW, Whyte LG. 2013. Bacterial growth at -15°C; molecular insights from the permafrost bacterium *Planococcus halocryophilus* Or1. ISME J 7:1211-1226.
21. Goordial J, Raymond-Bouchard I, Zolotarov Y, de Bethencourt L, Ronholm J, Shapiro N, Woyke T, Stromvik M, Greer CW, Bakermans C, Whyte L. 2016. Cold adaptive traits revealed by comparative genomic analysis of the eurypsychrophile *Rhodococcus* sp. JG3 isolated from high elevation McMurdo Dry Valley permafrost, Antarctica. FEMS Microbiol Ecol 92:fiv154.

22. Raymond-Bouchard I, Goordial J, Zolotarov Y, Ronholm J, Stromvik M, Bakermans C, Whyte LG. 2018. Conserved genomic and amino acid traits of cold adaptation in subzero-growing Arctic permafrost bacteria. FEMS Microbiol Ecol 94:fiy023.

23. Methé BA, Nelson KE, Deming JW, Momen B, Melamud E, Zhang XJ, Moult J, Madupu R, Nelson WC, Dodson RJ, Brinkac LM, Daugherty SC, Durkin AS, DeBoy RT, Kolonay JF, Sullivan SA, Zhou LW, Davidsen TM, Wu M, Huston AL, Lewis M, Weaver B, Weidman JF, Khouri H, Utterback TR, Feldblyum TV, Fraser CM. 2005. The psychrophilic lifestyle as revealed by the genome sequence of *Colwellia psychrerythraea* 34H through genomic and proteomic analyses. P Natl Acad Sci USA 102:10913-10918.

24. Dsouza M, Taylor MW, Turner SJ, Aislabie J. 2015. Genomic and phenotypic insights into the ecology of *Arthrobacter* from Antarctic soils. BMC Genomics 16.

25. Rabus R, Ruepp A, Frickey T, Rattei T, Fartmann B, Stark M, Bauer M, Zibat A, Lombardot T, Becker I. 2004. The genome of *Desulfotalea psychrophila*, a sulfate-reducing bacterium from permanently cold Arctic sediments. Environ Microbiol 6:887-902.

26. Feng S, Powell SM, Wilson R, Bowman JP. 2014. Extensive gene acquisition in the extremely psychrophilic bacterial species *Psychroflexus torquis* and the link to sea-ice ecosystem specialism. Genome Biol Evol 6:133-148.

27. Shen L, Liu Y, Xu B, Wang N, Zhao H, Liu X, Liu F. 2017. Comparative genomic analysis reveals the environmental impacts on two *Arcticibacter* strains including sixteen Sphingobacteriaceae species. Sci Rep 7:2055.

28. D'Amico S, Collins T, Marx JC, Feller G, Gerday C. 2006. Psychrophilic microorganisms: challenges for life. EMBO Rep 7:385-9.

29. Chen Z, Yu H, Li L, Hu S, Dong X. 2012. The genome and transcriptome of a newly described psychrophilic archaeon, *Methanolobus psychrophilus* R15, reveal its cold adaptive characteristics. Env Microbiol Rep 4:633-641.

30. Fournier GP, Andam CP, Gogarten JP. 2015. Ancient horizontal gene transfer and the last common ancestors. BMC Evol Biol 15:70.

31. Stokes L. 1963. Recent Progress in Microbiology, VIII. University of Toronto Press, Toronto.

32. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski AD, Pyshkin AV, Sirotkin AV, Vyahhi N, Tesler G, Alekseyev MA, Pevzner PA. 2012. SPAdes: A New Genome Assembly Algorithm and Its Applications to Single-Cell Sequencing. J Comput Biol 19:455-477.

33. Chen S, Zhou Y, Chen Y, Gu J. 2018. fastp: an ultra-fast all-in-one FASTQ preprocessor. Bioinformatics 34:i884-i890.

34. Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW. 2015. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. Genome Res 25:1043.

35. Lagesen K, Hallin P, Rodland EA, Staerfeldt H-H, Rognes T, Ussery DW. 2007. RNAmmer: consistent and rapid annotation of ribosomal RNA genes. Nucleic Acids Res 35:3100-3108.

36. Dee DP, Uppala SM, Simmons AJ, Berrisford P, Poli P, Kobayashi S, Andrae U, Balmaseda MA, Balsamo G, Bauer P. 2011. The ERA-Interim reanalysis: configuration and performance of the data assimilation system. Quartjroymeteorsoc 137:553-597.

37. Overbeek R, Olson R, Pusch GD, Olsen GJ, Davis JJ, Disz T, Edwards RA, Gerdes S, Parrello B, Shukla M, Vonstein V, Wattam AR, Xia FF, Stevens R. 2014. The SEED and the rapid

annotation of microbial genomes using subsystems technology (RAST). Nucleic Acids Res 42:D206-D214.

38. Vesth T, Lagesen K, Acar O, Ussery D. 2013. CMG-Biotools, a free workbench for basic comparative microbial genomics. PloS One 8:e60120.

39. Yarza P, Richter M, Peplies J, Euzeby J, Amann R, Schleifer KH, Ludwig W, Glöckner FO, Rosselló-Móra R. 2008. The All-Species Living Tree project: A 16S rRNA-based phylogenetic tree of all sequenced type strains. Syst Appl Microbiol 31:241-250.

40. Yang ZH. 2006. Computational Molecular Evolution. Oxford University Press, Great Britain.

41. Parks DH, Rinke C, Chuvochina M, Chaumeil PA, Woodcroft BJ, Evans PN, Hugenholtz P, Tyson GW. 2017. Recovery of nearly 8,000 metagenome-assembled genomes substantially expands the tree of life. Nat Microbiol 2:1533-1542.

42. Edgar RC. 2004. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. BMC Bioinformatics 5:1-19.

43. Castresana J. 2000. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. Mol Biol Evol 17:540-52.

44. Stamatakis A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 30:1312-1313.

45. Nguyen LT, Schmidt HA, von Haeseler A, Minh BQ. 2015. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol Biol Evol 32:268-274.

46. Ronquist F, Huelsenbeck JP. 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics 19:1572-1574.

47. Lanfear R, Calcott B, Ho SYW, Guindon S. 2012. PartitionFinder: combined selection of partitioning schemes and substitution models for phylogenetic analyses. Mol Biol Evol 29:1695-1701.

48. Letunic I, Bork P. 2016. Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. Nucleic Acids Res 44:W242-W245.

49. Dixon P. 2003. VEGAN, a package of R functions for community ecology. J Veg Sci 14:927-930.

50. Csuroes M. 2010. Count: evolutionary analysis of phylogenetic profiles with parsimony and likelihood. Bioinformatics 26:1910-1912.

51. Oliveira PH, Touchon M, Cury J, Rocha EPC. 2017. The chromosomal organization of horizontal gene transfer in bacteria. Nat Commun 8:841.

52. Orlandini V, Maida I, Fondi M, Perrin E, Papaleo MC, Bosi E, de Pascale D, Tutino ML, Michaud L, Lo Giudice A, Fani R. 2014. Genomic analysis of three sponge-associated *Arthrobacte*r Antarctic strains, inhibiting the growth of *Burkholderia cepacia* complex bacteria by synthesizing volatile organic compounds. Microbiol Res 169:593-601.

53. Nakatsu CH, Barabote R, Thompson S, Bruce D, Detter C, Brettin T, Han C, Beasley F, Chen W, Konopka A, Xie G. 2013. Complete genome sequence of *Arthrobacter* sp. strain FB24. Stand Genomic Sci 9:106-16.

54. Gkorezis P, Bottos EM, Van Hamme JD, Thijs S, Rineau F, Franzetti A, Balseiro-Romero M, Weyens N, Vangronsveld J. 2015. Draft genome sequence of *Arthrobacter* sp. strain SPG23, a hydrocarbon-degrading and plant growth-promoting soil bacterium. Genome Announc 3.

55. Bai Y, Muller DB, Srinivas G, Garrido-Oter R, Potthoff E, Rott M, Dombrowski N, Munch PC,

582 Spaepen S, Remus-Emsermann M, Huttel B, McHardy AC, Vorholt JA, Schulze-Lefert P.
583 2015. Functional overlap of the *Arabidopsis* leaf and root microbiota. Nature 528:364-9.
584 56. Shen L, Yao TD, Liu YQ, Jiao NZ, Kang SC, Xu BQ, Zhang SH. 2014. Downward-shifting
585 temperature range for the growth of snow-bacteria on glaciers of the Tibetan Plateau.
586 Geomicrobiol J 31:779-787.
587 57. Siddiqui KS, Cavicchioli R. 2006. Cold-adapted enzymes, p 403-433, Annual Review of
588 Biochemistry, vol 75.
589 58. Anderson MJ, Walsh DCI. 2013. PERMANOVA, ANOSIM, and the Mantel test in the face of
590 heterogeneous dispersions: What null hypothesis are you testing? Ecol Monogr 83:557-574.
591 59. Allen MA, Lauro FM, Williams TJ, Burg D, Siddiqui KS, De Francisci D, Chong KWY, Pilak
592 O, Chew HH, De Maere MZ, Ting L, Katrib M, Ng C, Sowers KR, Galperin MY, Anderson IJ,
593 Ivanova N, Dalin E, Martinez M, Lapidus A, Hauser L, Land M, Thomas T, Cavicchioli R.
594 2009. The genome sequence of the psychrophilic archaeon, *Methanococcoides burtonii*: the
595 role of genome evolution in cold adaptation. ISME J 3:1012-1035.
596 60. Richards VP, Palmer SR, Bitar PDP, Qin X, Weinstock GM, Highlander SK, Town CD, Burne
597 RA, Stanhope MJ. 2014. Phylogenomics and the dynamic genome evolution of the genus
598 *Streptococcus*. Genome Biol Evol 6:741-753.
599 61. Cuypers TD, Hogeweg P. 2012. Virtual genomes in flux: an interplay of neutrality and
600 adaptability explains genome expansion and streamlining. Genome Biol Evol 4:212-29.
601 62. Van de Guchte M. 2017. Horizontal gene transfer and ecosystem function dynamics. Trends
602 Microbiol 25:699-700.
603 63. Collins RE, Deming JW. 2013. An inter-order horizontal gene transfer event enables the
604 catabolism of compatible solutes by Colwellia psychrerythraea 34H. Extremophiles
605 17:601-610.
606 64. Nelson-Sathi S, Sousa FL, Roettger M, Lozada-Chavez N, Thiergart T, Janssen A, Bryant D,
607 Landan G, Schonheit P, Siebers B, McInerney JO, Martin WF. 2015. Origins of major archaeal
608 clades correspond to gene acquisitions from bacteria. Nature 517:77-80.
609 65. Mackelprang R, Burkert A, Haw M, Mahendrarajah T, Conaway CH, Douglas TA, Waldrop
610 MP. 2017. Microbial survival strategies in ancient permafrost: insights from metagenomics.
611 ISME J 11:2305-2318.
612 66. Demaere MZ, Williams TJ, Allen MA, Brown MV, Gibson JAE, Rich J, Lauro FM,
613 Dyallsmith M, Davenport KW, Woyke T. 2013. High level of intergenera gene exchange
614 shapes the evolution of haloarchaea in an isolated Antarctic lake. P Natl Acad Sci USA
615 110:16939.
616 67. Dziewit L, Bartosik D. 2014. Plasmids of psychrophilic and psychrotolerant bacteria and their
617 role in adaptation to cold environments. Front Microbiol 5:596.
618 68. Ma Y, Wang L, Shao Z. 2006. Pseudomonas, the dominant polycyclic aromatic
619 hydrocarbon-degrading bacteria isolated from Antarctic soils and the role of large plasmids in
620 horizontal gene transfer. Environ Microbiol 8:455-465.
621 69. Bahl J, Lau MC, Smith GJ, Vijaykrishna D, Cary SC, Lacap DC, Lee CK, Papke RT,
622 Warren-Rhodes KA, Wong FK, McKay CP, Pointing SB. 2011. Ancient origins determine
623 global biogeography of hot and cold desert cyanobacteria. Nat Commun 2:163.
624 70. Choudoir MJ, Buckley DH. 2018. Phylogenetic conservatism of thermal traits explains
625 dispersal limitation and genomic differentiation of *Streptomyces* sister-taxa. ISME J

12:2176-2186.

71. Zhang DC, Schumann P, Liu HC, Xin YH, Zhou YG, Schinner F, Margesin R. 2010. *Arthrobacter alpinus* sp. nov., a psychrophilic bacterium isolated from alpine soil. Int J Syst Evol Microbiol 60:2149-53.

72. Shen L, Liu Y, Gu Z, Yao T, Xu B, Wang N, Jiao N, Liu H, Zhou Y. 2015. *Arcticibacter eurypsychrophilus* sp. nov., isolated from ice core. Int J Syst Evol Microbiol 65:639-643.

73. Shen L, Liu Y, Yao T, Wang N, Xu B, Jiao N, Liu H, Zhou Y, Liu X, Wang Y. 2013. *Dyadobacter tibetensis* sp. nov., isolated from glacial ice core. Int J Syst Evol Microbiol 63:3636-3639.

74. Shen L, Liu Y, Gu Z, Xu B, Wang N, Jiao N, Liu H, Zhou Y. 2015. *Massilia eurypsychrophila* sp. nov. a facultatively psychrophilic bacteria isolated from ice core. Int J Syst Evol Microbiol 65:2124-2129.

75. Xing T, Yao T, Liu Y, Wang N, Xu B, Shen L, Gu Z, Gu B, Liu H, Zhou Y. 2016. *Polaromonas eurypsychrophila* sp. nov., isolated from an ice core. Int J Syst Evol Microbiol 66:2497-2501.

76. Shivaji S, Reddy GSN, Chattopadhyay MK. 2017. Bacterial Biodiversity, Cold Adaption and Biotechnological Importance of Bacteria Occurring in Antarctica. Proc Ind Natl Sci Acad 83:327-352.

77. Vincent WF. 2000. Evolutionary origins of Antarctic microbiota: invasion, selection and endemism. Antarct Sci 12:374-385.

78. Irgens RL, Gosink JJ, Staley JT. 1996. *Polaromonas vacuolata* gen. nov., sp. nov., a psychrophilic, marine, gas vacuolate bacterium from Antarctica. Int J Syst Bacteriol 46:822-6.

79. Ganzert L, Bajerski F, Mangelsdorf K, Lipski A, Wagner D. 2011. Leifsonia psychrotolerans sp nov., a psychrotolerant species of the family Microbacteriaceae from Livingston Island, Antarctica. International Journal of Systematic and Evolutionary Microbiology 61:1938-1943.

Figure legends

Figure 1 (a) Phylogenetic clustering of *Arthrobacter* strains based on concatenated alignment of 119 orthologous proteins using RAxML, numbers at nodes indicate posterior probability/bootstrap percentages by MrBayes and RAxML. (b) Relative composition (frequency compared to the total amino acids present) of lysine and arginine in protein sequences of *Arthrobacter* proteomes, arrows highlighted significant change in the relative composition of arginine and lysine. (c) Average annual air temperature of the strains (from 2008 to 2017), the environment

663 temperature of the polar and alpine strains was significantly lower than that of the
664 references strains ($P < 0.005$, F = 42.1).

665

666 Figure 2 Growth curves for the Group A (*A. luteolus*, *A. globiformis* and *A.*
667 *subterraneus*), group B (*Arthrobacter* sp. 4R501, *Arthrobacter* sp. 9E14 and
668 *Arthrobacter* sp. 08Y14), group C (*A. alpinus*, *Arthrobacter* sp. A3 and *Arthrobacter*
669 sp. N199823) grow at (a) 25 °C, (b) 5 °C and (c) -1 °C.

670

671 Figure 3 Changes in overall amino acid frequencies between (a) group A and B, (b)
672 group A and C.

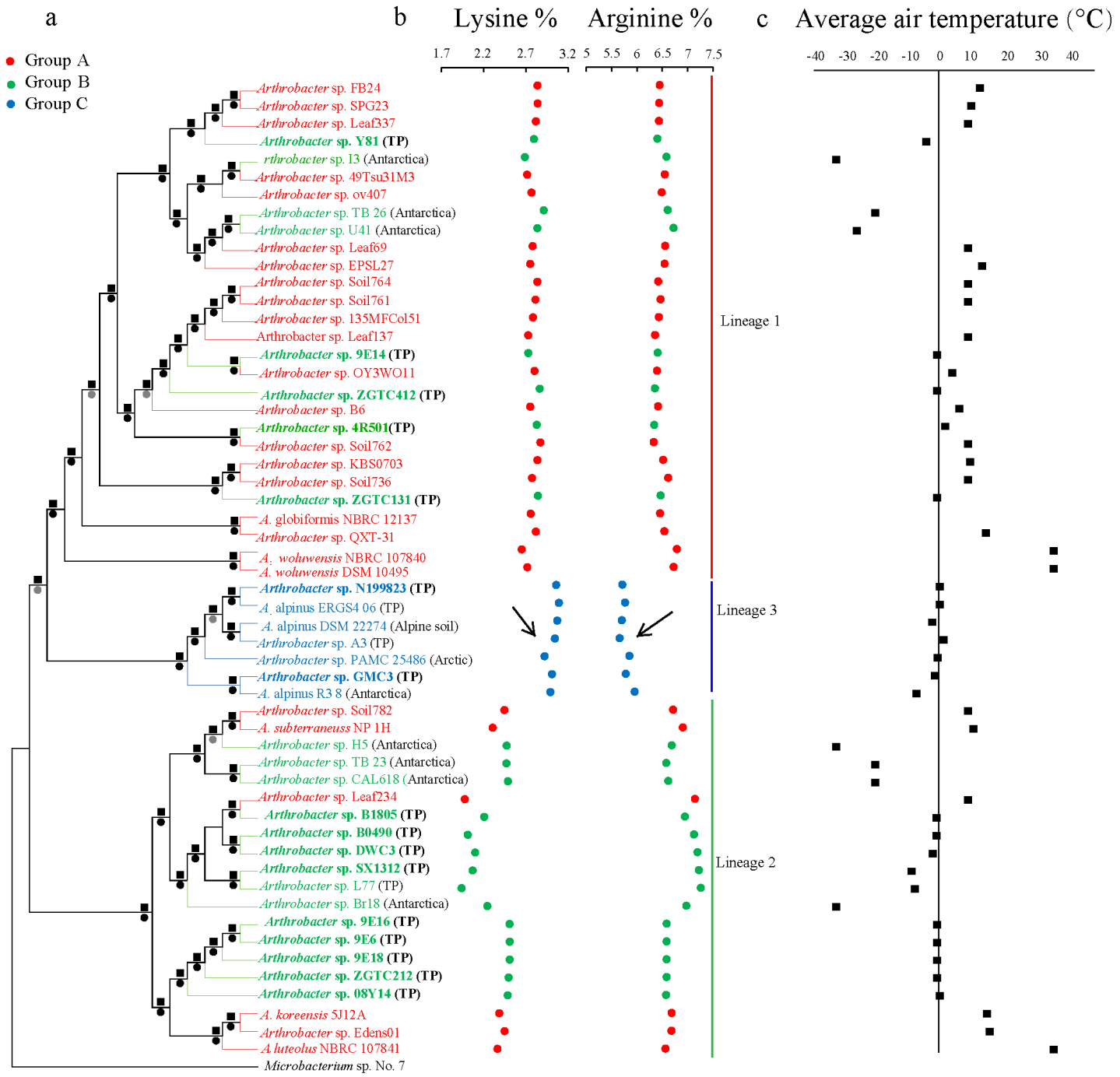
673

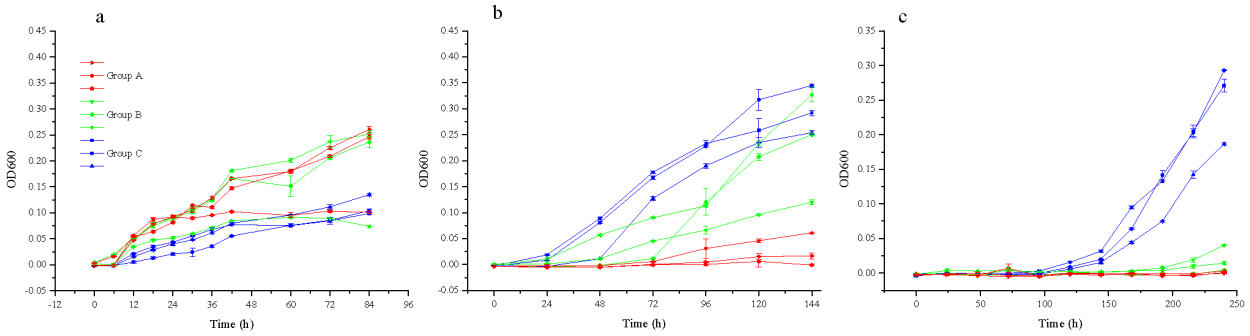
674 Figure 4 (a) Ordination of functional genes (classified at the level of function) using
675 two-dimensional nonmetric multidimensional scaling (NMDS). Shading shows clear
676 separation of group C (PERMANOVA, $P = 0.001$). (b) The frequency of significant
677 indicator genes ($P < 0.05$) in (1) a comparison of group A and C strains, (2) a
678 comparison of group A and B strains. (c) Phylogenetic relationships of genes related
679 to cold shock; a close clustering between *cspA* and *cspC* from genomes of group C
680 shows that these genes are monophyletic and can be clearly separated from those from
681 group A and B which have an interleaved clustering of genes.

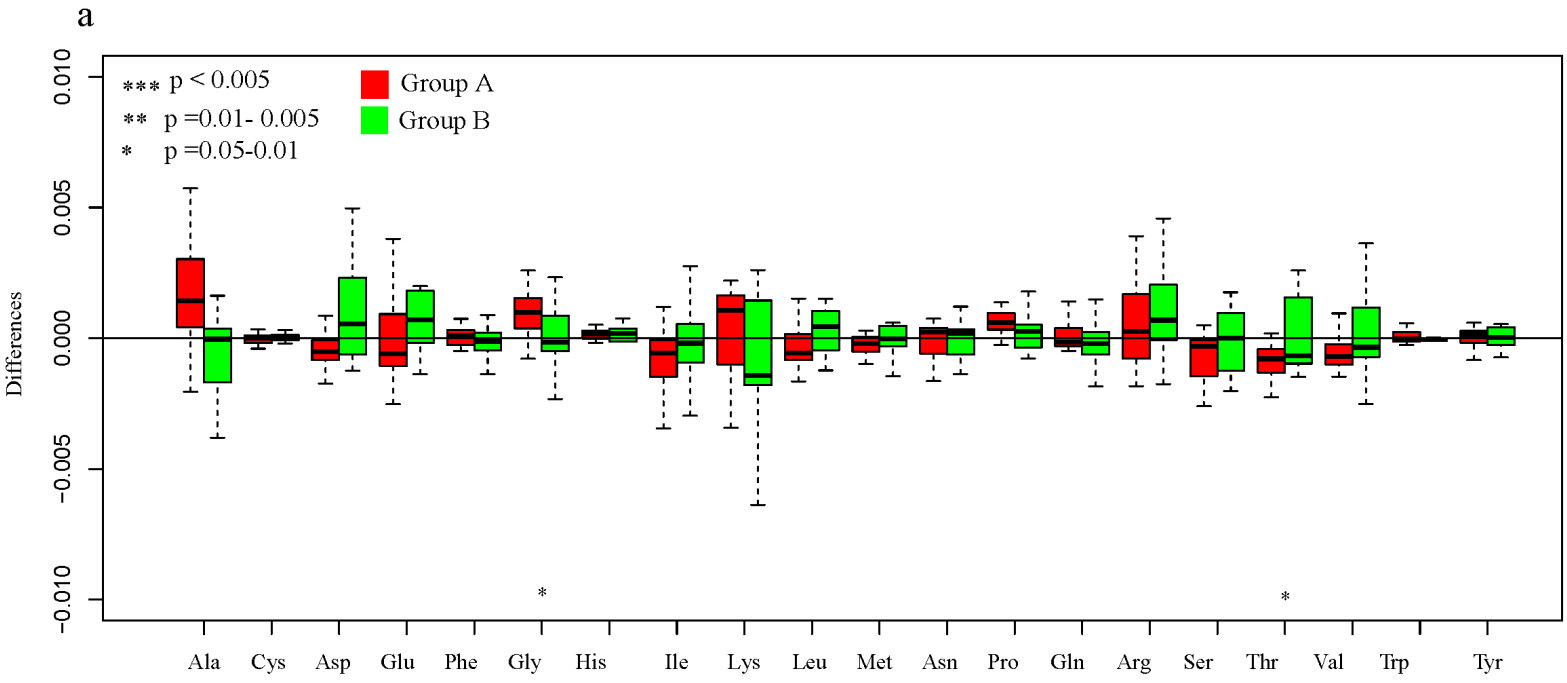
682

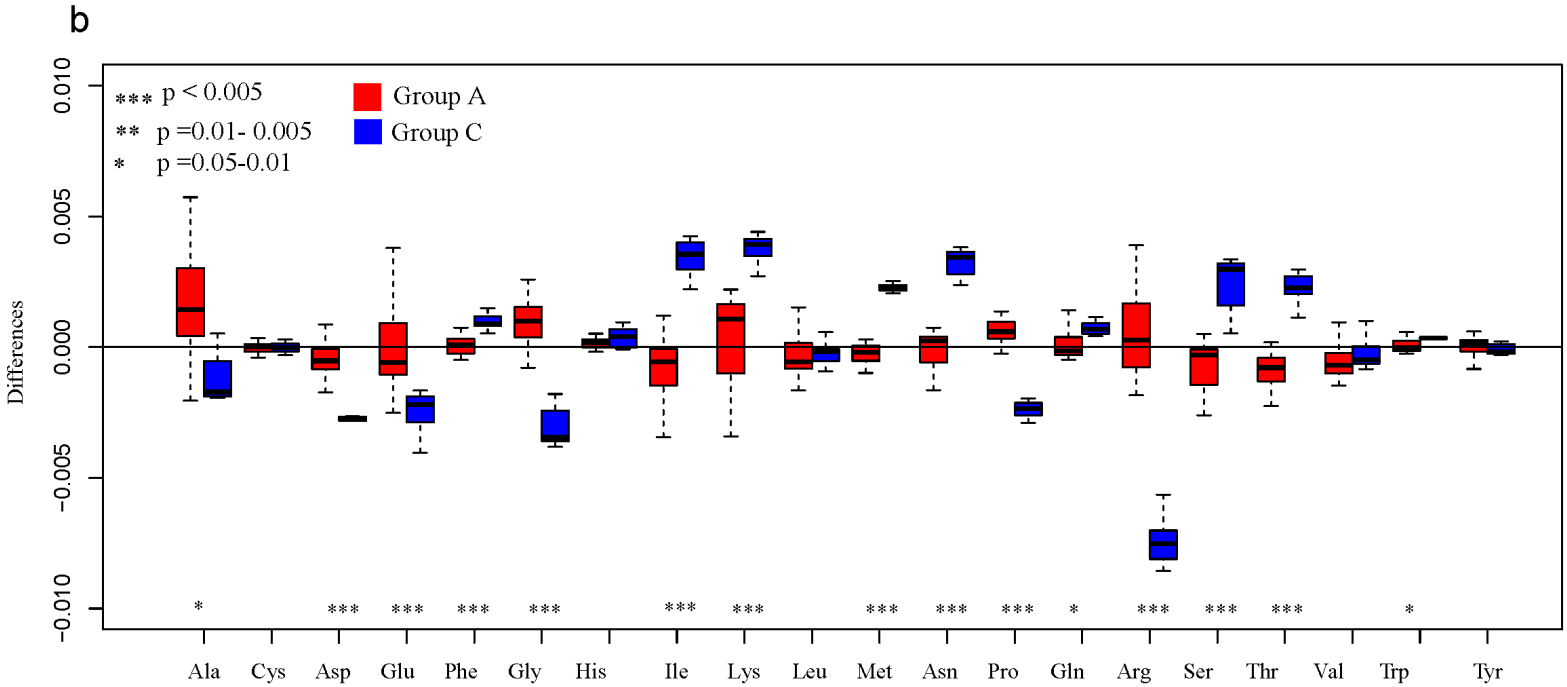
683 Figure 5 (a) Ancestral genome content reconstruction using COUNT software. The
684 reconstruction is based on the RAxML *Arthrobacter* tree. The log-scale color coding

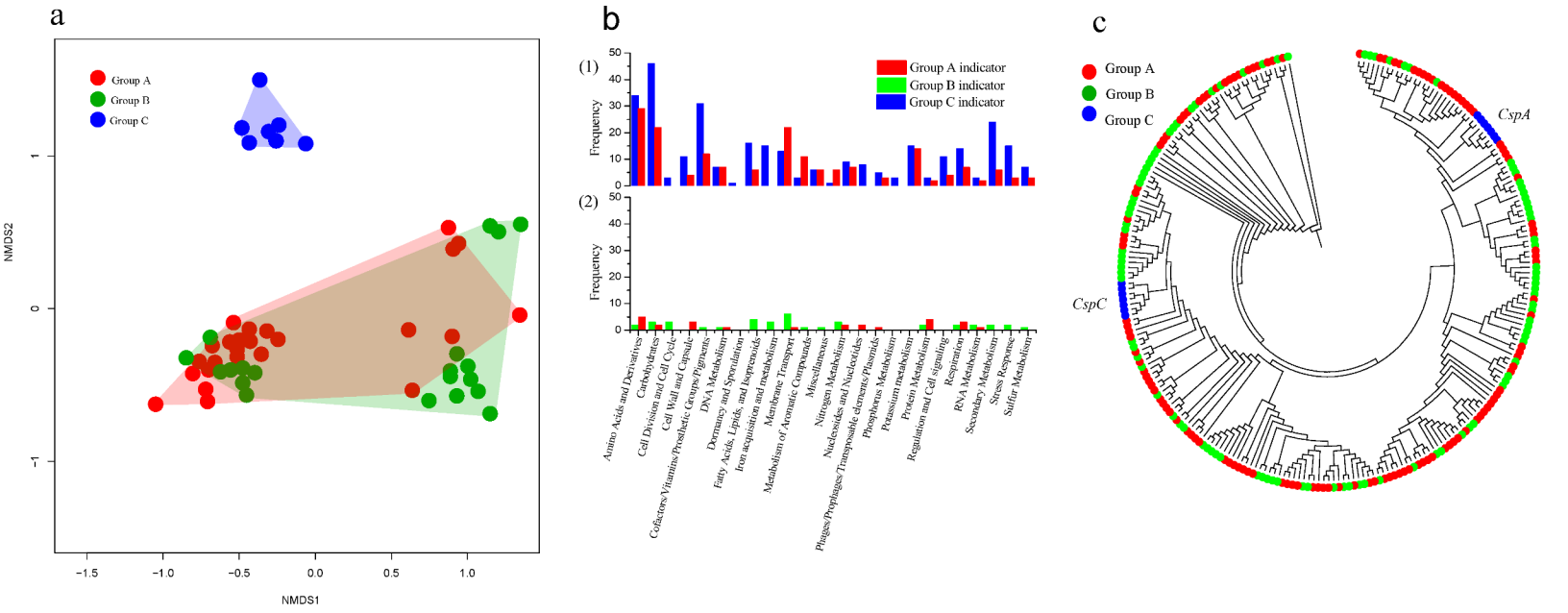
685 represents the number of reconstructed gain and loss events for each lineage.
686 Numbers in parentheses are predicted gene numbers for ancestral nodes and observed
687 gene numbers for extant lineages. (b) Condensed and linearized maximum likelihood
688 cladogram, showing the genome dynamics on the branches leading to nodes N27, N33
689 and N53. (c) Bar plot showing the distribution of genes families gained at N27, N33
690 and N53.
691

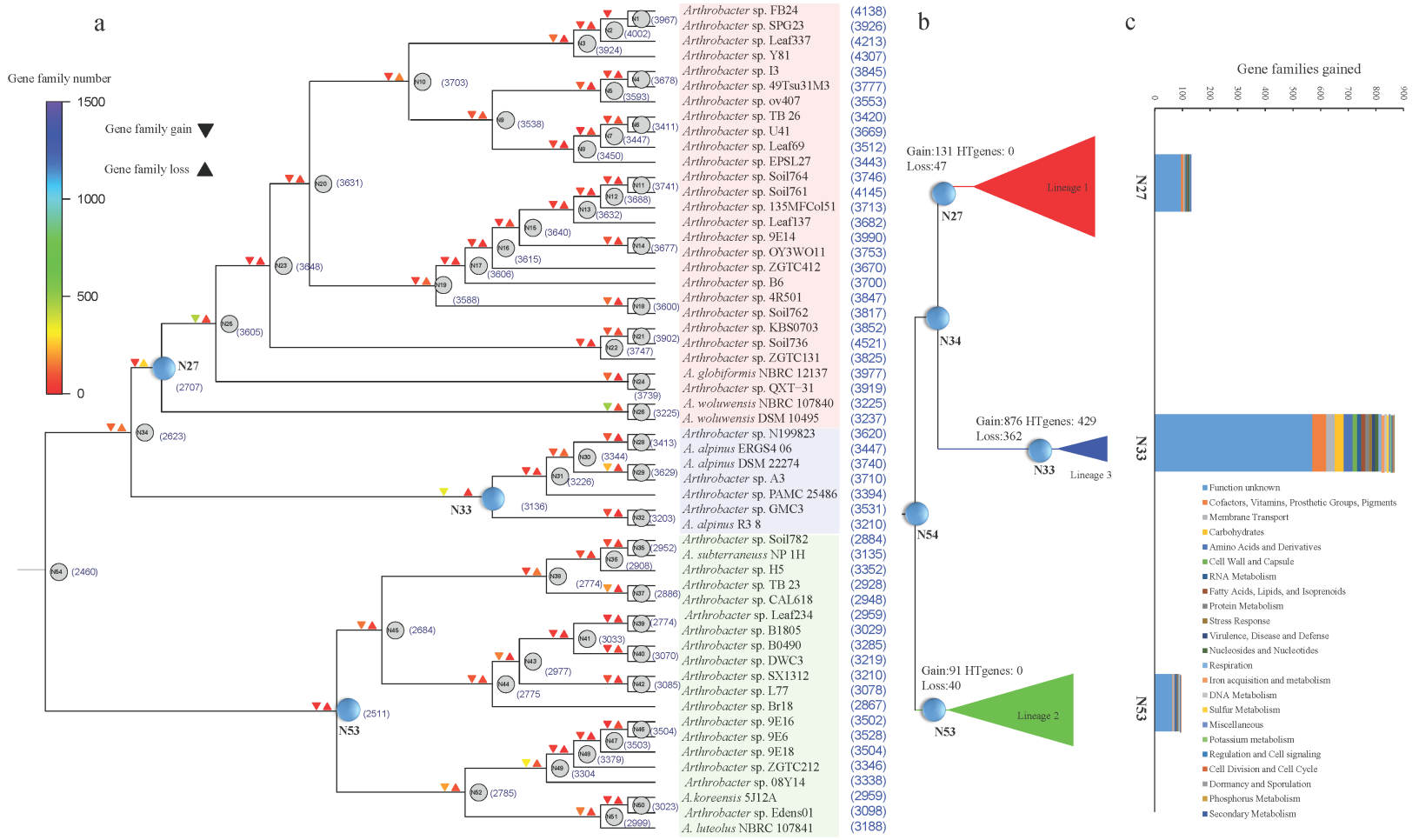

a
b
Lysine %
Arginine %
c
Average air temperature (°C)
Group A
Group B
Group C
1.7 2.2 2.7 3.2
5 5.5 6 6.5 7 7.5
-40 -30 -20 -10 0 10 20 30 40
Arthrobacter sp. FB24
Arthrobacter sp. SPG23
Arthrobacter sp. Leaf337
Arthrobacter sp. Y81 (TP)
rthrobacter sp. I3 (Antarctica)
Arthrobacter sp. 49Tsu31M3
Arthrobacter sp. ov407
Arthrobacter sp. TB 26 (Antarctica)
Arthrobacter sp. U41 (Antarctica)
Arthrobacter sp. Leaf69
Arthrobacter sp. EPSL27
Arthrobacter sp. Soil764
Arthrobacter sp. Soil761
Arthrobacter sp. 135MFCol51
Arthrobacter sp. Leaf137
Arthrobacter sp. 9E14 (TP)
Arthrobacter sp. OY3WO11
Arthrobacter sp. ZGTC412 (TP)
Arthrobacter sp. B6
Arthrobacter sp. 4R501(TP)
Arthrobacter sp. Soil762
Arthrobacter sp. KBS0703
Arthrobacter sp. Soil736
Arthrobacter sp. ZGTC131 (TP)
A. globiformis NBRC 12137
Arthrobacter sp. QXT-31
A. woluwensis NBRC 107840
A. woluwensis DSM 10495
Arthrobacter sp. N199823 (TP)
A. alpinus ERGS4 06 (TP)
A. alpinus DSM 22274 (Alpine soil)
Arthrobacter sp. A3 (TP)
Arthrobacter sp. PAMC 25486 (Arctic)
Arthrobacter sp. GMC3 (TP)
A. alpinus R3 8 (Antarctica)
Arthrobacter sp. Soil782
A. subterraneuss NP 1H
Arthrobacter sp. H5 (Antarctica)
Arthrobacter sp. TB 23 (Antarctica)
Arthrobacter sp. CAL618 (Antarctica)
Arthrobacter sp. Leaf234
Arthrobacter sp. B1805 (TP)
Arthrobacter sp. B0490 (TP)
Arthrobacter sp. DWC3 (TP)
Arthrobacter sp. SX1312 (TP)
Arthrobacter sp. L77 (TP)
Arthrobacter sp. Br18 (Antarctica)
Arthrobacter sp. 9E16 (TP)
Arthrobacter sp. 9E6 (TP)
Arthrobacter sp. 9E18 (TP)
Arthrobacter sp. ZGTC212 (TP)
Arthrobacter sp. 08Y14 (TP)
A. koreensis 5J12A
Arthrobacter sp. Edens01
A luteolus NBRC 107841
Microbacterium sp. No. 7
Lineage 1
Lineage 3
Lineage 2


a
b
c
OD600
Time (h)
Group A
Group B
Group C


a
*** p < 0.005
** p =0.01- 0.005
* p =0.05-0.01
Group A
Group B
Differences
0.010
0.005
0.000
−0.005
−0.010
Ala Cys Asp Glu Phe Gly His Ile Lys Leu Met Asn Pro Gln Arg Ser Thr Val Trp Tyr


b
*** p < 0.005
** p =0.01- 0.005
* p =0.05-0.01
Group A
Group C
Differences
0.010
0.005
0.000
−0.005
−0.010
Ala Cys Asp Glu Phe Gly His Ile Lys Leu Met Asn Pro Gln Arg Ser Thr Val Trp Tyr


a
Group A
Group B
Group C
NMDS1
NMDS2
b
(1)
(2)
Frequency
Group A indicator
Group B indicator
Group C indicator
Amino Acids and Derivatives
Carbohydrates
Cell Division and Cell Cycle
Cell Wall and Capsule
Cofactors,Vitamins,Prosthetic Groups,Pigments
DNA Metabolism
Dormancy and Sporulation
Fatty Acids, Lipids, and Isoprenoids
Iron acquisition and metabolism
Membrane Transport
Metabolism of Aromatic Compounds
Miscellaneous
Nitrogen Metabolism
Nucleosides and Nucleotides
Phages,Prophages,Transposable elements,Plasmids
Phosphorus Metabolism
Potassium metabolism
Protein Metabolism
Regulation and Cell signaling
Respiration
RNA Metabolism
Secondary Metabolism
Stress Response
Sulfur Metabolism
c
Group A
Group B
Group C
CspA
CspC


a
Gene family number
1500
1000
500
0
Gene family gain
Gene family loss
N27
(2707)
N33
(3136)
N53
(2511)
N54
(2460)
Arthrobacter sp. FB24 (4138)
Arthrobacter sp. SPG23 (3926)
Arthrobacter sp. Leaf337 (4213)
Arthrobacter sp. Y81 (4307)
Arthrobacter sp. I3 (3845)
Arthrobacter sp. 49Tsu31M3 (3777)
Arthrobacter sp. ov407 (3553)
Arthrobacter sp. TB 26 (3420)
Arthrobacter sp. U41 (3669)
Arthrobacter sp. Leaf69 (3512)
Arthrobacter sp. EPSL27 (3443)
Arthrobacter sp. Soil764 (3746)
Arthrobacter sp. Soil761 (4145)
Arthrobacter sp. 135MFCol51 (3713)
Arthrobacter sp. Leaf137 (3682)
Arthrobacter sp. 9E14 (3990)
Arthrobacter sp. OY3WO11 (3753)
Arthrobacter sp. ZGTC412 (3670)
Arthrobacter sp. B6 (3700)
Arthrobacter sp. 4R501 (3847)
Arthrobacter sp. Soil762 (3817)
Arthrobacter sp. KBS0703 (3852)
Arthrobacter sp. Soil736 (4521)
Arthrobacter sp. ZGTC131 (3825)
A. globiformis NBRC 12137 (3977)
Arthrobacter sp. QXT−31 (3919)
A. woluwensis NBRC 107840 (3225)
A. woluwensis DSM 10495 (3237)
Arthrobacter sp. N199823 (3620)
A. alpinus ERGS4 06 (3447)
A. alpinus DSM 22274 (3740)
Arthrobacter sp. A3 (3710)
Arthrobacter sp. PAMC 25486 (3394)
Arthrobacter sp. GMC3 (3531)
A. alpinus R3 8 (3210)
Arthrobacter sp. Soil782 (2884)
A. subterraneus NP 1H (3135)
Arthrobacter sp. H5 (3352)
Arthrobacter sp. TB 23 (2928)
Arthrobacter sp. CAL618 (2948)
Arthrobacter sp. Leaf234 (2959)
Arthrobacter sp. B1805 (3029)
Arthrobacter sp. B0490 (3285)
Arthrobacter sp. DWC3 (3219)
Arthrobacter sp. SX1312 (3210)
Arthrobacter sp. L77 (3078)
Arthrobacter sp. Br18 (2867)
Arthrobacter sp. 9E16 (3502)
Arthrobacter sp. 9E6 (3528)
Arthrobacter sp. 9E18 (3504)
Arthrobacter sp. ZGTC212 (3346)
Arthrobacter sp. 08Y14 (3338)
A.koreensis 5J12A (2959)
Arthrobacter sp. Edens01 (3098)
A. luteolus NBRC 107841 (3188)
b
Gain:131 HTgenes: 0
Loss:47
Lineage 1
N27
N34
Gain:876 HTgenes: 429
Loss:362
N33
Lineage 3
N54
Gain:91 HTgenes: 0
Loss:40
Lineage 2
N53
c
Gene families gained
0
100
200
300
400
500
600
700
800
900
N27
N33
N53
Function unknown
Cofactors, Vitamins, Prosthetic Groups, Pigments
Membrane Transport
Carbohydrates
Amino Acids and Derivatives
Cell Wall and Capsule
RNA Metabolism
Fatty Acids, Lipids, and Isoprenoids
Protein Metabolism
Stress Response
Virulence, Disease and Defense
Nucleosides and Nucleotides
Respiration
Iron acquisition and metabolism
DNA Metabolism
Sulfur Metabolism
Miscellaneous
Potassium metabolism
Regulation and Cell signaling
Cell Division and Cell Cycle
Dormancy and Sporulation
Phosphorus Metabolism
Secondary Metabolism