Comparative genome analysis reveals the molecular basis of nicotine degradation and survival capacities of Arthrobacter

Yuxiang Yao1,2*, Hongzhi Tang1,2*, Fei Su1,2 & Ping Xu1,2

1State Key Laboratory of Microbial Metabolism, and School of Life Sciences & Biotechnology, Shanghai Jiao Tong University, Shanghai 200240, People’s Republic of China, 2Joint International Research Laboratory of Metabolic & Developmental Sciences, Shanghai Jiao Tong University, Shanghai 200240, People’s Republic of China.

Arthrobacter is one of the most prevalent genera of nicotine-degrading bacteria; however, studies of nicotine degradation in Arthrobacter species remain at the plasmid level (plasmid pAO1). Here, we report the bioinformatic analysis of a nicotine-degrading Arthrobacter aurescens M2012083, and show that the moeB and mogA genes that are essential for nicotine degradation in Arthrobacter are absent from plasmid pAO1. Homologues of all the nicotine degradation-related genes of plasmid pAO1 were found to be located on a 68,622-bp DNA segment (nic segment-1) in the M2012083 genome, showing 98.1% nucleotide acid sequence identity to the 69,252-bp nic segment of plasmid pAO1. However, the rest sequence of plasmid pAO1 other than the nic segment shows no significant similarity to the genome sequence of strain M2012083. Taken together, our data suggest that the nicotine degradation-related genes of strain M2012083 are located on the chromosome or a plasmid other than pAO1. Based on the genomic sequence comparison of strain M2012083 and six other Arthrobacter strains, we have identified 17 transcription factors reported to be involved in stress responses and 109 genes involved in environmental adaptability of strain M2012083. These results reveal the molecular basis of nicotine degradation and survival capacities of Arthrobacter species.

Arthrobacter species, which were first isolated from soil in 18891, are high GC-content bacteria, and typically appear as gram-negative rods in early cultivation stages, but as gram-positive cocci in later stages2. Strains of these species are among the most common aerobic culturable bacteria in soil and are thought to act as the predominant decomposers of organic matter3. Arthrobacter species have been reported to be able to degrade environmental pollutants such as nitroglycerin, many benzene derivatives, polycyclic aromatic compounds, haloalcohols, haloalkanes, N-heterocyclic compounds, insecticides, and herbicides4–19. They have been found in extreme environments such as arctic ice, the deep subsurface, and radioactive environments; in addition, chemically contaminated and heavy metal-contaminated sites seem to be rich in these bacteria20–24. The ubiquity of Arthrobacter species is probably due to their tolerance to various stresses such as long-term starvation, desiccation, oxidative stress, temperature shifts, osmotic pressure changes, ionizing radiation, excess heavy metal ions, and toxic chemicals22–29. Their remarkable survival abilities contribute to the importance of Arthrobacter species in pollutant degradation in complex and volatile environments27. Therefore, in order to reveal the application potential of a pollutant-degrading bacterium in environmental remediation, it is necessary to understand not only the genes involved in pollutant degradation but also the genes involved in environmental survivability.

Because of the widespread presence and high toxicity of N-heterocyclic pollutants, their degradation has received great attention30. Some N-heterocyclic pollutant-degrading Arthrobacter species such as the s-triazine-degrading strain A. aurescens TC1 (TC1) and quinaldine-degrading strain Arthrobacter nitroguajacolicus Rue61a (Rue61a) have been isolated from atrazine-containing soil and the sewage sludge of a coal tar refining factory, respectively18,31,32. Furthermore, their complete genomes have been sequenced, and the molecular bases of their metabolic and survival potentials have been thoroughly analyzed27,33.
The complete genome sequences of four other *Arthrobacter* species including three soil strains, *Arthrobacter* sp. FB24 (FB24), *Arthrobacter chlorophenolicus* A6 (A6), and *Arthrobacter phenanthrenovorans* Sphe3 (Sphe3), and a food-processing strain, *Arthrobacter arilaitensis* Re117 (Re117), were reported. Strains FB24, A6 and Sphe3 were isolated by their ability to degrade dimethylbenzene, 4-chlorophenol, and phenanthrene, respectively. Genome analyses revealed many metabolic and stress-related genes in these environmental strains, which reflected their pollutant-degrading abilities and niche specializations. Strain Re117 was isolated from the surface of Reblochon cheese; its genetic adaptation to the habitat via genes such as those involved in iron acquisition, salt tolerance, and catabolism of fatty acids, amino acids, and lactic acid, which are major carbon substrates present at the cheese surface, was also revealed by genome sequence analysis.

Although nicotine is a typical N-heterocyclic pollutant and *Arthrobacter* is one of the most prevalent genera of nicotine-degrading bacteria, studies of nicotine degradation by *Arthrobacter* species remain at the plasmid level, while the complete genome of another prevalent bacterium, *Pseudomonas putida* S16, has been sequenced and extensively studied. A catabolic plasmid of 165-kb isolated from *Arthrobacter nicotinovorans* (formerly classified as *Arthrobacter oxydans*) was sequenced, and almost all the genes involved in nicotine degradation by the bacterium were shown to locate on it. Two related genes that were absent from the plasmid are required for molybdenum cofactor (MoCo) biosynthesis; one encodes Moeb, which is essential for ATP-dependent activation of the MoAD subunit of molybdopterin (MPT) synthase; the other encodes a MoaB/MgoA homolog, which is probably required for MPT-adenylate formation. These genes must locate on the chromosome, and genome sequence analysis will help us to establish this. Furthermore, the survival capacity of nicotine-degrading *Arthrobacter* strains, which is very important to their application potential, is still unknown. Therefore, it is necessary to obtain and analyze the genome sequence of a nicotine-degrading *Arthrobacter* strain in order to uncover the molecular basis of its survival capacity.

Many catabolic plasmids are thought to take part in the spread of catabolic traits by horizontal gene transfer between bacteria. Recently, Mihasan and Brandsch proposed that the nicotine-degrading genes of *A. nicotinovorans* were possibly obtained via plasmid pAO1 or via a precursor plasmid from the chromosome of a strain that was probably related to *Rhodococcus opacus* or another *Arthrobacter* species. However, *Arthrobacter* species harboring nicotine-degrading genes on the chromosome have not yet been reported.

Recently, we isolated a new nicotine-degrading strain, *Arthrobacter* sp. M2012083 (M2012083), from tobacco waste, and obtained and published its genome sequence, but without further analysis. In the present study, we classified this strain as *Arthrobacter aurescens* by physiological and biochemical identification, 16S rDNA phylogenetic analysis, and genomic analysis; we compared the genomic sequence of strain M2012083 with the six available *Arthrobacter* complete genomes and the sequence of nicotine-degrading plasmid pAO1, in order to provide a better understanding of the genetic basis of the nicotine-degradation and survival capacities of this genus.

**Results**

**Classification of strain M2012083.** The nicotine-degrading strain M2012083 was isolated from tobacco waste and deposited in the China Center for Type Culture Collection (CCTCC; collection number M2012083). Strain M2012083 is gram-positive, aerobic, and asporous. Transmission electron microscopic observation revealed that the cells of strain M2012083 were long rods in their lag phase (6 h) and became coccus-shaped during the log phase (36 h) in nicotine-containing medium. These morphological characteristics are representative of the genus *Arthrobacter*, but were also detected in members of other genera such as *Brevibacterium* and *Rhodococcus*. The results of carbon source oxidation tests, enzyme activity tests, and carbon source utilization tests of strain M2012083 are shown in Table S1, Table S2, and Table S3, respectively. The 16S rDNA sequence of strain M2012083 was sequenced and submitted to GenBank accession number KF893303. Nucleotide BLAST searches of the GenBank database and Ribosomal Database Project library (http://blast.ncbi.nlm.nih.gov/Blast.cgi) indicated that the 16S rDNA sequence of strain M2012083 exhibited more than 99% nucleotide acid sequence identities with that of TC1 and Rue61a. We selected 55 16S rDNA sequences from different strains in order to construct a phylogenetic tree by using MEAG5 analysis with the neighbor-joining method. Strain M2012083 is closest in phylogenetic tree to type strains of *A. aurescens* and *A. nitroguajaciclus* (Fig. 1). Furthermore, genome comparison using the RAST Prokaryotic Genome Annotation Server showed that the genomic sequence of M2012083 was more similar to the complete genome of TC1 (comparison score: 546) than to that of Rue61a (comparison score: 524) (Fig. 2). Based on the above results, strain M2012083 was classified as *A. aurescens*.

**General genome features and comparative genomics.** The general features of the seven subject *Arthrobacter* genomes are summarized in Table 1. Pan-genome analyses, such as an overview of genomic conserved regions, orthologous group categorization, and orthologous relationship analysis, were performed in order to identify conserved and strain-specific CDSs. BLASTAtlas tool was used to compare the genomes of the seven *Arthrobacter* strains and to provide a quick overview of the conserved genomic regions. The genome of TC1 was used as a reference because it is most similar to the genome of M2012083 and has been thoroughly studied. In the seven different *Arthrobacter* genomes, 68–79% of protein-coding genes were functionally categorized based on the Clusters of Orthologous Groups (COGs) database. Orthologous relationship analysis was performed using the OrthoMCL method. All protein-coding genes of the seven subject strains were clustered into a total of 5,368 orthologous groups with 1,653 conserved groups, while the protein-coding genes of M2012083 were clustered into a total of 3,774 groups (Dataset1, Fig. 5). The strain-specific CDSs of each of the seven *Arthrobacter* strains, which have no homologous genes in the other six strains, are listed in Dataset2, along with their ORF IDs and annotations.

**Nicotine catabolism.** In order to reveal the nicotine-degradation mechanism of strain M2012083, its genome was searched for all reported bacterial nicotine degradation-related genes and the protein-coding genes of plasmid pAO1, using the tBLASTp program. In the M2012083 genome, there are 94 of the total 166 ORFs of plasmid pAO1, including all 52 nicotine degradation-related ORFs and some carbohydrate catabolism-related ORFs (Fig. 6, Dataset3).

**Survival capacity of *Arthrobacter* strains.** Genome analysis and comparisons showed that more than 100 ORFs related to survival capacity occurred in all subject *Arthrobacter* strains except Re117, which contained only 74 ORFs (Table 2). ORFs involved in survival capacity are listed in detail for all subject *Arthrobacter* strains in Dataset4.

**Discussion**

The assembled genome of M2012083 had approximately 329-fold sequence coverage, with 4,312 protein-coding genes. The genome size of the seven subject *Arthrobacter* strains ranges from 3.8 to 5.3 Mbp. The smallest genome is that of Re117, which is the only food-processing bacterium among the subject strains, resulting in the fewest protein-coding genes (Table 1); this suggests that living on the surface of cheese requires the fewest genes. To a certain extent,
Figure 1 | 16S rDNA phylogenetic analysis of *Arthrobacter* strains. The evolutionary history was inferred using the Neighbor-Joining method\(^5\). The optimal tree with the sum of branch length = 0.82 is shown. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1000 replicates) are shown next to the branches\(^7\). The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. The evolutionary distances were computed using the Neighbour-Joining (NJ) method\(^7\) and are in the units of the number of base substitutions per site. The analysis involved 55 nucleotide sequences. All positions containing gaps and missing data were eliminated. There were a total of 1,327 positions in the final DataSet. Evolutionary analyses were conducted in MEGA5\(^1\). Only bootstrap values greater than 50% are shown. The strains used in genome comparison are shown in red, the type strains are indicated by superscript T and the strain M2012083 is highlighted in blue. The accession numbers of the sequences used are listed in the parentheses.
which are close on the phylogenetic tree, show similar GC content, genetic relationship, and function. For example, the Arthrobacter strains can provide data on the links between GC content, genetic relationship, and function. For example, the Arthrobacter strains were submitted to the web service RAST for comparison. The genome of M2012083 was compared with the other six complete genomes and the comparison scores were shown. Higher score means higher similarity.

The different genomic G + C content of the seven studied Arthrobacter strains can provide data on the links between G + C content, genetic relationship, and function. For example, the Arthrobacter strains were submitted to the web service RAST for comparison. The genome of M2012083 was compared with the other six complete genomes and the comparison scores were shown. Higher score means higher similarity.

![Figure 2](image) Genome similarity analysis of strain M2012083. The genomes of the seven Arthrobacter strains were submitted to the web service RAST for comparison. The genome of M2012083 was compared with the other six complete genomes and the comparison scores were shown. Higher score means higher similarity.

### Table 1 | General genome features of the seven subject Arthrobacter strains

| Genome feature | M2012083 | FB24 | TC1 | A6 | Re17 | Sphe3 | Rue61a |
|----------------|---------|-----|-----|----|------|-------|--------|
| Size (base pair) | 4,629,172 | 5,070,478 | 5,226,648 | 4,980,870 | 3,918,192 | 4,535,320 | 5,081,038 |
| G + C content | 62.0% | 65.4% | 62.4% | 66.0% | 59.3% | 65.4% | 62.2% |
| Protein-coding genes | 4312 | 4523 | 4588 | 4590 | 3436 | 4131 | 4575 |
| No. of protein-coding genes with function prediction | 3254 (75.5%) | 3256 (72.0%) | 3366 (73.4%) | 3095 (67.4%) | 2378 (69.2%) | 2922 (70.7%) | 3168 (69.2%) |
| No. of protein-coding genes without function prediction | 1058 (24.5%) | 1267 (28.0%) | 1222 (26.6%) | 1495 (32.6%) | 1058 (30.8%) | 1209 (29.3%) | 1407 (30.8%) |
| RNA genes | 3114 (72.2%) | 3567 (78.9%) | 3361 (73.3%) | 3125 (68.1%) | 2473 (72.0%) | 3264 (79.0%) | 3367 (73.6%) |
| rRNA genes | 58 | 69 | 73 | 104 | 82 | 65 | 71 |
| (5S rRNA, 16S rRNA, 23S rRNA) | (2, 1, 1) | (5, 5, 5) | (6, 6, 6) | (5, 5, 5) | (6, 6, 6) | (4, 4, 4) | (6, 6, 6) |
| tRNA genes | 54 | 51 | 54 | 88 | 64 | 50 | 53 |
| Other RNA genes | 0 | 3 | 1 | 0 | 0 | 0 | 0 |
| Conserved CDS | 1694 | 1717 | 1722 | 1703 | 1677 | 1688 | 1704 |
| Strain-specific CDS | 434 | 540 | 427 | 788 | 816 | 417 | 328 |

The orthologous relationship analysis was performed by OrthoMCL. a: Protein-coding genes having orthologous genes in every other subject genomes; b: Protein-coding genes not having orthologous gene in any other subject genomes.

All the nicotine degradation-related genes, such as those encoding nicotine dehydrogenase subunits (ndhL, ndhM, and ndhS), 6-hydroxy-L-nicotine oxidase (6hlno), 6-hydroxy-D-nicotine oxidase (6hdmno), and keto dehydrogenase subunits (kdhL, kdhM, and kdhS) were found in the Re117 genome. The numbers of genes categorized as COG C, K, R, S, and T in the Re117 genome are also much lower than those in other subject genomes, which indicates that surviving on the surface of cheese requires less functions related to energy transformation, transcription, or signal transduction. The COGs have highly similar distributions among the six environmental Arthrobacter strains, except for COGs L and N. Genes with COG L annotations, representing replication, recombination, and repair, are more numerous in Sphe3 than in the other five strains. There are 15 genes annotated as COG N in A6, which is far more than that in other strains (only 0 or 1), suggesting that all the subject Arthrobacter strains lacked cell motility, except strain A6. Genes annotated as COG B, the group related to chromatin structure and dynamics, were not found in any of the seven subject genomes (Fig. 4). Comparisons between the M2012083 genome and each of the other six subject genomes indicated that 3,575 orthologous groups in M2012083 were conserved in Rue61a, while 199 groups were specific to M2012083 compared with Rue61a. In contrast, M2012083 and Re117 only share 2,004 conserved groups, and 1,770 groups in M2012083 are specific to M2012083 compared with Re117. There were 3,369 conserved groups among the genomes of the N-heterocyclic pollutant-degrading bacteria, M2012083, TC1, and Rue61a, including 356 groups that were not found in the other four subject genomes. Among the six environmental Arthrobacter strains, M2012083, FB24, TC1, A6, Sphe3, and Rue61a, 2,323 conserved groups were identified, including 670 groups that were not found in the food-processing strain, Re117 (Fig. 5).
kdhS), and those involved in γ-N-methylaminobutyrate catabolism (mgaba), MoCo biosynthesis (moco), assembly and quality control of the (αβγδ)2 holoenzyme complexes (aqc), compound transport (perm1 and perm2), and gene regulation (tr1, tr2, and tr3), were located on a 68,622-bp DNA segment (nic segment-1) of contig 058 in the M2012083 genome, which showed 98.1% nucleotide acid sequence identity to the 69,252-bp nic segment of pAO1 (Fig. 6, DataSet3). Furthermore, the only nicotine degradation-related ORFs absent from plasmid pAO1, the two which encode the MoeB protein and a MogA homolog that are essential for MoCo biosynthesis were also identified in the M2012083 genome (1197706.3.peg.1697 and 1197706.3.peg.1089). However, the xerD integrase gene and the IS1473 insertion sequence that contains the resolvase gene are absent from the nic segment-1 of M2012083, while the sequence of the Tn554 transposon is incomplete within it (Fig. 6, DataSet3). Plasmid pAO1 sequences other than the nic segment, which contain all the plasmid-function genes, show no significant similarity to the M2012083 genome sequences (Fig. 6). Most of the

Figure 3 | Comparison of the genome of Arthrobacter aurescens M2012083 with other Arthrobacter genomes. The outermost circle (circle 1) represents the scale. Circle 2, the chromosomal (blue) and plasmids (orange and green) open reading frames (ORFs) of A. aurescens TC1 as references; circle 3, ORFs of A. chlorphenolicus A6 complete genome; circle 4, ORFs of A. nitroguajacolicus Rue61a complete genome; circle 5, ORFs of A. arilaitensis Re117 complete genome; circle 6, ORFs of A. phenanthrenivorans Sphe3 complete genome; circle 7, ORFs of A. sp. FB24 complete genome; circle 8, ORFs of A. aurescens M2012083 complete genome. ORFs are represented by colorful sticks (red-to-blue were assigned according to the similarity of the ORF to the homolog in TC1 genome) in circles 3 to 8.
pAO1 ORFs involved in plasmid function, such as the ORFs involved in replication (CDS_28 and CDS_120), partitioning (PAR_1 and CDS_9), maintenance (CDS_21 and CDS_22), conjugation (CDS_5, CDS_6, CDS_83, and CDS_92), stress response (CDS_7), abortive infection phage resistance (CDS_118), cell-cell contact (CDS_18), and cell signaling (CDS_113), and many predicted membrane proteins that may participate in conjugation (CDS_8, CDS_14, CDS_30, CDS_31, CDS_84, CDS_85, CDS_86, CDS_87, CDS_88, CDS_89, CDS_90, CDS_91, and CDS_114), were not found in the M2012083 genome (DataSet3). These data suggest that the nicotine degradation-related genes of *A. aurescens* M2012083 are located on the chromosome or a plasmid other than pAO1.

It has been reported that alternative σ-factor induction is an important strategy for bacteria to deal with environmental stress. The genome of M2012083 encodes 35 transcription factors; these include 17 members of the σ^70 family of σ factors, this number is...
similar to that in the other two N-heterocyclic pollutant-degrading Arthrobacter strains (TC1 and Rue61a) and considerably more than that in strain Re117, which only encodes six (Table 2). The expression of universal stress-related proteins (USPs) is induced in cells in response to heat shock, oxidant and UV exposure, carbon, nitrogen, and phosphate starvation, and entering stationary phase. As in strain TC1, M2012083 contains eight potential USPs, far fewer than FB24 and Rue61a (15 and 13, respectively) but more than Re117.

| Arthrobacter strain | Transcription factors | α20 factors | USPs | Osmotic stress | Cold and heat shock | Detoxification | Other stress | Totala |
|---------------------|----------------------|-------------|------|----------------|-------------------|--------------|------------|--------|
| M2012083            | 35                   | 17          | 8    | 35             | 29                | 21           | 4          | 12     | 109    |
| FB24                | 29                   | 13          | 17   | 33             | 31                | 18           | 3          | 12     | 111    |
| TC1                 | 34                   | 17          | 8    | 32             | 30                | 19           | 3          | 12     | 104    |
| A6                  | 28                   | 12          | 11   | 34             | 30                | 20           | 6          | 11     | 112    |
| Re117               | 22                   | 6           | 6    | 20             | 26                | 15           | 3          | 4      | 74     |
| Sphe3               | 27                   | 10          | 11   | 30             | 32                | 19           | 6          | 10     | 108    |
| Rue61a              | 35                   | 18          | 13   | 35             | 31                | 20           | 3          | 12     | 114    |

- Total number of ORFs related to USPs, osmotic stress, oxidative stress, cold and heat shock, detoxification and other stress.
Selenium is essential for living organisms, at low concentration, and occurs extensively in the Earth’s crust. However, in aerobic environments, selenium occurs as selenate and selenite, which are toxic and mutagenic at excess concentration. The dedA gene is reported to be very important for the regulation of selenite uptake in a high selenite-resistant bacterium, *Ralstonia metallidurans* CH34. Three dedA genes were identified in strain M2012083 and two in FB24, TC1, A6, Sphe3, and Rue61a, while only one dedA gene was identified in Re117. However, only the ATP-binding component of the selenite uptake transporter is present in all the subject strains, while the periplasmic substrate-binding and permease components of the ABC transporter exist only in strains A6 and Sphe3 (DataSet4).

In addition to σ^2^- factors, σ^3^ and RpoS family members, which are alternative σ factors for RNA polymerase transcription, play major roles in the general stress response and environmental survival of bacteria. σ^2^- activity is inhibited by anti-σ^2^- factor RsbW, while it is stimulated by serine phosphatase RsbU and anti-σ^2^- factor antagonist RsbV. Strains M2012083, FB24, TC1, A6, and Rue61a contain genes that encode RNA polymerase σ factor SigB, RsbU, and RsbV, while strain Sphe3 contains genes that encode SigB and RsbU. In contrast to other subject *Arthrobacter* strains, strain Re117 does not contain any genes involved in σ^2^- stress-response regulation. The rbsW gene was not detected in any of the subject strains (DataSet4).

During the response to carbon starvation, rpoS gene expression is inhibited by RpsA, which is a starvation-sensing protein. The rpsA gene and a carbon starvation protein A gene were identified in strain M2012083 and the other environmental subject strains. In contrast, genes involved in the carbon starvation response were not detected in strain Re117 (DataSet4). Expression of the rpoS gene is also regulated by homoserine lactones (HSLs) and a derivative thereof. Genes encoding aspartate kinase, aspartate semialdehyde dehydrogenase, and homoserine dehydrogenase, which has been suggested to catalyze HSL synthesis from α-aspartate, were identified in the M2012083 genome (1197706.3.peg.3479, 1197706.3.peg.1452, and 1197706.3.peg.2912, respectively) and the other six *Arthrobacter* genomes (DataSet1). Based on all the genome analyses, strain Re117 contains the fewest carbohydrate metabolism and stress-response genes of the seven subject *Arthrobacter* strains; this is probably because the habitat of Re117, the surface of cheese, is much more sterile than soil.

**Methods**

**Strain identification.** The M2012083 strain was cultured at 30°C and 220 rpm in lysogenic broth or nicotine medium, as previously described. Morphological characteristics were observed using a transmission electron microscope. Physiological and biochemical properties, such as the oxidation and utilization of different carbon sources, and the activities of different enzymes were determined by the CCTCC. The 16S rDNA of strain M2012083 was amplified by PCR with the universal primer pair of 27F (5’-AGAGTTTGATCCTGGCTCAG-3’) and 1492R (5’-GGTTACCTTGTTACGACTT-3’) from extracted genomic DNA. The PCR product was purified and sequenced, then analyzed by homology alignment using the BLAST search program (http://www.ncbi.nlm.nih.gov/BLAST.html). A phylogenetic tree was constructed by the neighbor-joining method using MEGAS analysis.

**Genome sequencing and annotation.** The genomic DNA of M2012083 was extracted using the Wizard Genomic DNA Purification Kit (Promega, Madison, WI, USA). The genome of strain M2012083 was sequenced by the Chinese National Human Genome Center (Shanghai, China). The Velvet program was used to assemble the pair-end reads de novo with manually optimized settings. The genome sequences of FB24, TC1, A6, Re117, Sphe3, and Rue61a were obtained from GenBank. The genomes of all these *Arthrobacter* strains were submitted to the RAST web service for automated annotation followed by manual checking. The annotations are available via a guest account at the RAST website. Information on COGs and from the Conserved Domain Database was also used for comparisons.
Genome comparison and analysis. The National Institutes of Allergy and Infectious Diseases/Pathosystems Resource Integration Center (PATRIC) and RAST were used for genome comparisons and metabolic pathway analyses. The TBLASTn program used was to search for homologs of plasmid pAO1 protein-coding genes in the M2012083 genome with coding amino acid sequence identity > 30% at an e-value < 1e-5. The MUMmer program was used to generate a synteny plot of plasmid pAO1 versus the M2012083 genome. The OrthoMCL method was used to analyze the orthologous relationships between protein-coding genes in different genomes. Genes in different genomes with coding amino acid sequence identity > 50% at an e-value < 1e-5 were identified as orthologous genes. We used the CIRSPRFinder program in CIRSPR’s web server to identify clustered regularly interspaced short palindromic repeats in the CIRSPR/Cas RNA-guided nucleic system. The genomic context was visualized using the Circos software.

Accession numbers. The genome sequences of Arthrobacter aurescens M2012083 has been deposited in NCBI database under the accession number AKKK00000000. The accession numbers of the subject genomes are as following: Arthrobacter sp. FB24 chromosome, NC_008541; Arthrobacter sp. FB24 plasmid 1, NC_008537; Arthrobacter sp. FB24 plasmid 2, NC_008538; Arthrobacter aurescens TC1 chromosome, NC_008711; Arthrobacter aurescens TC1 plasmid TC1, NC_008712; Arthrobacter aurescens TC1 plasmid TC2, NC_008713; Arthrobacter chlorophenolicus A6 chromosome, NC_011886; Arthrobacter chlorophenolicus A6 plasmid pACH01, NC_011879; Arthrobacter chlorophenolicus A6 plasmid pACH02, NC_011881; Arthrobacter arilaitensis Re117 chromosome, NC_014550; Arthrobacter arilaitensis Re117 plasmid pRE117-1, NC_014549; Arthrobacter arilaitensis Re117 plasmid pRE117-2, NC_014548; Arthrobacter phanenothreoprinos Re8a3 chromosone, NC_015145; Arthrobacter phanenothreoprinos Re8a3 plasmid pAPH301, NC_015146; Arthrobacter phanenothreoprinos Re8a3 plasmid pAPH302, NC_015147; Arthrobacter nitroguaajacilus Rue61a chromosome, NC_018531; Arthrobacter nitroguaajacilus Rue61a plasmid pRue61a_0032, NC_018532; Arthrobacter nitrocitrinivorans plasmid pAO1, NC_012229.
Yao, Y. et al. Genome sequence of a nicotine-degrading strain of *Arthrobacter*. *J. Bacteriol.* **194**, 5714–5715 (2012).

Holt, J. G., Sneath, P., Mair, M. & Sharpe, P. *Bergey's manual of systematic bacteriology*. 8th edn (Williams and Wilkins press, 1986).

Tamura, K. et al. MEGAS: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol. Biol. Evol.* **28**, 2731–2739 (2011).

Saitou, N. & Nei, M. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* **4**, 406–425 (1987).

Hallin, P. F., Binniewies, T. T. & Ussery, D. W. The genome BLASTatlas—a GeneWiz extension for visualization of whole-genome homology. *Mol. Biosyst.* **4**, 363–371 (2008).

Tatusov, R. L. et al. The COG database: new developments in phylogenetic classification of proteins from complete genomes. *Nucleic Acids Res.* **29**, 22–28 (2001).

Li, L., Stoeckert, C. J. & Roos, D. S. OrthoMCL: identification of ortholog groups for eukaryotic genomes. *Genome Res.* **13**, 2178–2189 (2003).

Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–410 (1990).

Qiu, J. et al. A sirA-like gene, *sirA2*, is essential for 3-succinyl-pyridine metabolism in the newly isolated nicotine-degrading *Pseudomonas* sp. HZN6 strain. *Appl. Microbiol. Biotechnol.* **92**, 1023–1032 (2011).

Qiu, J. et al. Functional identification of two novel genes from *Pseudomonas* sp. strain HZN6 involved in the catabolism of nicotine. *Appl. Environ. Microbiol.* **78**, 2154–2160 (2012).

Qiu, J., Ma, Y., Zhang, J., Wen, Y. & Liu, W. Cloning of a novel nicotine oxidase gene from *Pseudomonas* sp. strain HZN6 whose product nonenantioselectively degrades nicotine to pseudoxy nicot ine. *Appl. Environ. Microbiol.* **79**, 2164–2171 (2013).

Kvint, K., Nachin, L., Diez, A. & Nyström, T. The bacterial universal stress protein: function and regulation. *Curr. Opin. Microbiol.* **6**, 140–145 (2003).

Potts, M. Desiccation tolerance of prokaryotes. *Microbiol. Rev.* **58**, 755–805 (1994).

Zevenhuizen, L. Levels of trehalose and glycin in *Arthrobacter globiformis* under conditions of nutrient starvation and osmotic stress. *Antonie Van Leeuwenhoek* **61**, 61–68 (1992).

Boscari, A., Mandon, K., Poggi, M.-C. & Le Rudulier, D. Functional expression of *Sinorhizobium meliloti BetS*, a high-affinity betaine transporter, in *Bradyrhizobium japonicum* USDA110. *Appl. Environ. Microbiol.* **70**, 5916–5922 (2004).

Hungria, M. et al. Tolerance to stress and environmental adaptability of *Chromobacterium violaceum*. *Genet. Mol. Res.* **10**, 102–116 (2011).

Ledgham, F., Quest, B., Vallaeys, T., Mercay, M. & Covès, J. A probable link between the DedA protein and resistance to selenium. *Res. Microbiol.* **156**, 367–374 (2005).

 Bhattacharya et al. Comparative genome analysis reveals the molecular basis of nicotine degradation and survival capacities of *Arthrobacter*. *Sci. Rep.* **5**, 8642; DOI:10.1038/srep08642 (2015).