Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Comparative genomics of the metal tolerant soil bacterium Achromobacter strain H380

Song Liu
Key Laboratory of Animal Embryo Engineering and Molecular Breeding of Hubei Province, Institute of Animal Sciences and Veterinary Medicine, Hubei Academy of Agricultural Sciences

Fan Chen
Key Laboratory of Animal Embryo Engineering and Molecular Breeding of Hubei Province, Institute of Animal Sciences and Veterinary Medicine, Hubei Academy of Agricultural Sciences

Hao Gu
Key Laboratory of Animal Embryo Engineering and Molecular Breeding of Hubei Province, Institute of Animal Sciences and Veterinary Medicine, Hubei Academy of Agricultural Sciences

Changfan Zhou
Key Laboratory of Animal Embryo Engineering and Molecular Breeding of Hubei Province, Institute of Animal Sciences and Veterinary Medicine, Hubei Academy of Agricultural Sciences

Wei Zeng
Key Laboratory of Animal Embryo Engineering and Molecular Breeding of Hubei Province, Institute of Animal Sciences and Veterinary Medicine, Hubei Academy of Agricultural Sciences

Hongwei Xiao
Key Laboratory of Animal Embryo Engineering and Molecular Breeding of Hubei Province, Institute of Animal Sciences and Veterinary Medicine, Hubei Academy of Agricultural Sciences

Yanzhen Bi
Key Laboratory of Animal Embryo Engineering and Molecular Breeding of Hubei Province, Institute of Animal Sciences and Veterinary Medicine, Hubei Academy of Agricultural Sciences

Hongyan Ren
renhy@hbaas.ac.cn

Key Laboratory of Animal Embryo Engineering and Molecular Breeding of Hubei Province, Institute of Animal Sciences and Veterinary Medicine, Hubei Academy of Agricultural Sciences

---

Research Article

Keywords: Achromobacter, comparative genomics, metal tolerance, phylogenomics, pan-genome, environmental adaptation

Posted Date: November 11th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11419809/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Comparative genomics of the metal tolerant soil bacterium *Achromobacter* strain H380

Song Liu[1], Fan Chen[1], Hao Gu[1], Changfan Zhou[1], Wei Zeng[1], Hongwei Xiao[1], Yanzhen Bi[1,2], Hongyan Ren[1*]

[1] *Key Laboratory of Animal Embryo Engineering and Molecular Breeding of Hubei Province, Institute of Animal Sciences and Veterinary Medicine, Hubei Academy of Agricultural Sciences, Wuhan 430070, China.*

[2] *Hubei Hongshan Laboratory, Wuhan 430070, China.*

*Correspondence to:

renhy@hbaas.ac.cn (Hongyan Ren).

Abstract

Background

Metal-tolerant environmental bacteria provide a tractable system for examining how genomic diversity relates to persistence under chemical stress. *Achromobacter* sp. H380, isolated from mining-affected soil, displays tolerance to several metals. Its genomic relationship to other *Achromobacter* strains and the distribution of its associated gene repertoire remain to be resolved.

Results

We combined the phenotypic profile of H380 with whole-genome sequencing and comparative analysis of 35 *Achromobacter* genomes. The H380 assembly comprised a 5,675,223-bp circular chromosome with a GC content of 67.94% and 5,278 predicted genes. A phylogeny based on 2,082 single-copy orthologous groups placed H380 in a seven-genome subclade. Within this subclade, H380 and ANB-1 shared 99.26% average nucleotide identity, whereas H380 showed substantially lower identity to the remaining strains. The seven-genome subclade contained 3,275 core and 6,979 pan-genome gene families, and H380 carried 461 strain-specific gene families within the 35-genome comparison. Seven subclade-associated candidates encoded predicted transport, regulatory and metabolic functions. Searches of environmental protein sequences recovered more distant homologues rather than close matches suitable for tracking the H380 subclade.

Conclusions

H380 combines a metal-tolerant phenotype with a distinct accessory gene repertoire within a genomically heterogeneous *Achromobacter* subclade. The analysis defines its comparative genomic context and identifies candidate functions for investigating environmental persistence and metal tolerance.

Keywords: *Achromobacter*; comparative genomics; metal tolerance; phylogenomics; pan-genome; environmental adaptation

1. Background

Metal contamination imposes persistent selection on soil microorganisms. Bacterial survival under these conditions depends on cellular processes that regulate metal uptake, export, sequestration and damage responses. Identifying tolerant isolates and defining their genomic context are therefore complementary steps towards understanding microbial persistence in contaminated environments [1, 2]. For bioremediation research, the immediate question is which traits sustain growth under exposure and which genomic features provide candidates for explaining those traits.

Achromobacter species occupy diverse environmental and host-associated habitats. Environmental isolates have been investigated for arsenite oxidation, hydrocarbon degradation and other transformations relevant to contaminated sites [3, 4, 5]. The genus also includes opportunistic pathogens, making strain-level characterization important when considering environmental applications. Genus-wide comparative work has linked core and accessory genomic components to host adaptation, while recent phenotypic and genomic analyses have documented substantial variation in antimicrobial susceptibility among *Achromobacter* species [6, 7]. These findings motivate joint consideration of phylogenetic placement, gene content and measured phenotype.

Achromobacter sp. H380 was isolated from mining-affected soil and deposited as CCTCC M2015291. Its use in cadmium-pollution treatment was described in an earlier patent [8]. The genomic basis of this strain's relationship to other *Achromobacter* isolates, however, requires a comparative framework that distinguishes shared ancestry from strain-specific variation. In particular, a phylogenetic subclade may contain several genomic species, and conserved functional annotations do not by themselves identify the genes responsible for a tolerance phenotype.

Here, we examine H380 using whole-genome sequencing and a comparison with 34 publicly available *Achromobacter* genomes. We first place its existing tolerance profile alongside the genome assembly, then resolve its relationships through phylogenomics and genome similarity. We subsequently compare gene repertoires, genome organization and predicted protein associations, and examine environmental homologues of selected subclade-associated genes. This framework connects an experimentally characterized environmental isolate to a defined comparative genomic context and a focused set of candidate functions.

2. Methods

2.1 Strain and phenotypic characterization

Achromobacter sp. H380 originated from mining-affected soil and was deposited in the China Center for Type Culture Collection, Wuhan, China, under accession CCTCC M2015291. Colony morphology, microscopy, growth during cadmium exposure and growth in metal- or antibiotic-containing media were recorded in the phenotypic overview (Fig. 1). The cadmium time course followed optical density at 600 nm over 72 h. The overview also reports metal-tolerance values and qualitative antibiotic-growth results in LB and M9 media.

2.2 Genome sequencing and annotation

Genomic DNA was extracted from H380. Single-molecule real-time sequencing was performed on a Pacific Biosciences RS platform by Novogene, Beijing, China. The sequencing run generated 86,786 reads with a mean length of 10,689 bp and a read N50 of 14,589 bp. De novo assembly was performed using the HGAP3 workflow, as described in the original analysis record, within the reported SMRT software environment [9].

The assembly was annotated using the NCBI Prokaryotic Genome Annotation Pipeline, reported as version 3.2 [10]. Ribosomal RNA and transfer RNA annotations were obtained with RNAmmer and ARAGORN, respectively; genomic islands were predicted with IslandPath-DIMOB [11, 12]. Predicted genes were assigned functional annotations using the COG, GO, KEGG and NCBI non-redundant protein databases [13, 14, 15].

2.3 Comparative dataset and phylogenetic analysis

The comparison comprised H380 and 34 *Achromobacter* assemblies downloaded from NCBI GenBank (Table 1). The species labels and assembly versions in Table 1 identify the dataset used in the original analysis. Orthologous gene families were identified with GET_HOMOLOGUES and the OrthoMCL algorithm [16, 17]. Single-copy orthologues shared across the 35 genomes were selected for phylogenetic reconstruction. Sequences were aligned using MUSCLE v3.6 and trimmed with the gappyout option in trimAl v1.4 [18, 19]. The alignments were concatenated, and maximum-likelihood inference was performed in IQ-TREE v1.5.5 with model selection based on the Bayesian information criterion [20, 21]. The reported selected model was GTR+F+R3. Trees were displayed using FigTree v1.4.3.

A Neighbor-Net network was constructed in SplitsTree v4.14.4 to examine relationships beyond a strictly bifurcating tree, with 1,000 bootstrap replicates reported for the network analysis [22, 23].

2.4 Genome similarity and gene-order comparisons

Average nucleotide identity (ANI) was calculated with OrthoANIu using USEARCH v8.1.1861 [24]. Average amino acid identity (AAI) was calculated using CompareM. Pairwise results were examined for the seven-genome subclade containing H380, ANB-1, LMG 3411, AD_173, PheN1, PR1 and MT3. Genome comparisons were visualized with the BLAST Ring Image Generator (BRIG), displaying the reported 80%, 90% and 100% identity bands [25]. MCScanX was used to identify collinear gene blocks and display gene-order relationships as dot plots and circular links [26].

2.5 Core and accessory gene repertoires

Core and pan-genome analyses used GET_HOMOLOGUES with the OrthoMCL clustering method and a reported minimum query coverage of 75%. Gene-family matrices were generated with the pipeline's matrix-processing utilities. Analyses were performed for the H380-containing subclade (n = 7), the remaining genomes (n = 28), and the complete dataset (n = 35). Core and pan-genome accumulation curves were generated using the pipeline's plotting utility and fitted using the reported Tettelin-based approach [27, 28]. Strain-specific families were defined by occurrence in only one genome within the analyzed dataset. Functional categories were assigned with WebMGA using RPS-BLAST against COG at an E-value threshold of 0.001 [29].

2.6 Predicted protein associations and environmental homology searches

Core protein sets from the three genome groups were queried against STRING v10.0 using *Achromobacter xylosoxidans* A8 as the reference organism [30]. Networks were visualized in Cytoscape v3.7.2, and connectivity was examined using NetworkAnalyzer and cytoHubba [31]. These networks represent predicted functional associations.

Genes present in the seven-genome subclade and absent from the other 28 genomes were identified from the gene-family matrix. Protein sequences were aligned with MUSCLE, trimmed with trimAl and compared using SIAS. Seven annotated candidates are listed in Table 2. Their encoded proteins were searched against the NCBI environmental protein database env_nr using BLASTP, and the reported best-hit coverage and identity were summarized (Table 3).

2.7 Reporting and use of AI assistance

Genomic comparisons are reported descriptively. No new significance tests or numerical estimates were derived from the plotted growth curves during manuscript preparation. The definitions of phenotypic replicates and error bars must be established from the experimental records before inferential analyses are reported. OpenAI Codex (OpenAI) assisted with manuscript

restructuring, English editing, consistency checking and document preparation. The authors are responsible for verification of the text, references, figures and interpretations.

3. Results

3.1 H380 displays a metal-tolerant phenotype

The H380 overview documents colony growth, rod-shaped cells and growth during cadmium exposure (Fig. 1). Across the displayed 72-h time course, optical density increased under all plotted cadmium conditions, with lower values at higher exposures. The accompanying metal table reports growth tolerance across cadmium, manganese, copper, zinc, chromium and cobalt conditions, with medium-dependent values for several metals. The antibiotic panel records growth at the displayed test concentrations in both LB and M9. These observations establish the experimental phenotype motivating the genomic analysis; the single-concentration antibiotic-growth panel is not a standardized clinical susceptibility classification.

3.2 The H380 genome contains diverse transport and metabolic functions

The assembly comprised a circular chromosome of 5,675,223 bp, with 67.94% GC content and 5,278 predicted genes. Annotation also identified 57 tRNA genes and eight predicted genomic islands. COG assignments included energy production and conversion, amino acid transport and metabolism, transcription, and inorganic ion transport and metabolism (Fig. 2A). GO annotations covered biological processes, cellular components and molecular functions (Fig. 2B). KEGG assignments included membrane transport and multiple metabolic pathways (Fig. 2C). These annotations define candidate cellular functions relevant to nutrient acquisition and environmental persistence.

The NR comparison assigned the largest number of protein matches to *Achromobacter* sequences, including 3,177 matches reported under *Achromobacter* sp. (Fig. 2D). This distribution provides annotation context; the phylogenomic and whole-genome comparisons below resolve relationships more directly than the frequency of database labels.

3.3 Phylogenomics identifies a seven-genome subclade containing H380

The maximum-likelihood analysis used 2,082 single-copy orthologous groups shared by the 35 genomes. H380 clustered closely with ANB-1 within a broader subclade also containing LMG 3411, AD_173, PheN1, PR1 and MT3 (Fig. 3A). The Neighbor-Net analysis recovered the same broad grouping (Fig. 3B). The dataset therefore places H380 in a defined lineage that contains strains carrying several species labels in the downloaded records.

Pairwise similarity resolved pronounced structure within this subclade (Fig. 4A,B). H380 and ANB-1 shared 99.26% ANI and 99.08% AAI. In contrast, H380 showed approximately 89–91% ANI to the other five members. AD_173, PheN1, PR1 and MT3 formed a second highly similar group, with pairwise ANI values above 99%. LMG 3411 had ANI values below 95% to the other subclade members. Thus, the seven-genome phylogenetic grouping contains distinct genome-similarity groups rather than establishing synonymy among all seven strains.

3.4 Genome comparisons reveal heterogeneous sequence conservation and gene order

The BRIG comparison showed uneven sequence conservation among the seven genomes relative to the reference (Fig. 4C). Dot plots and circular collinearity plots identified conserved gene blocks together with changes in their relative orientation and arrangement (Fig. 4D,E). The original analysis reported 25,611 collinear genes among 42,914 genes considered, corresponding to 59.68%.

H380, PheN1 and PR1 displayed prominent aligned blocks in the plotted coordinate system, whereas ANB-1 showed a different orientation pattern. Because the comparison includes both complete and fragmented assemblies (Table 1), these visual patterns describe genome organization at the resolution of the supplied assemblies. Reference orientation, contig ordering and assembly completeness must be distinguished from biological rearrangement when interpreting individual blocks.

3.5 Core and accessory gene repertoires differ across phylogenetic scales

The seven-genome subclade contained 3,275 core gene families and a pan-genome of 6,979 families (Fig. 5A,B). The remaining 28 genomes contained 2,546 core and 18,406 pan-genome families (Fig. 5C,D). Analysis of all 35 genomes further expanded the pan-genome and reduced the shared core relative to the 28-genome comparison (Fig. 5E,F). These patterns describe the combined effects of sampling breadth and phylogenetic diversity; they do not isolate a lineage-specific rate of gene gain or loss.

Strain-specific family counts ranged from 121 in AB2 to 1,014 in ANB-1, with 461 in H380 (Fig. 5G). The complete dataset also contained 13,049 families reported as shared by at least two genomes but outside the universal core. This accessory component supplies a basis for selecting H380-associated candidates beyond broadly conserved cellular functions.

3.6 Functional profiles separate conserved cellular processes from candidate lineage-associated functions

Core and accessory gene annotations were dominated by broad categories including amino acid transport and metabolism, energy production and conversion, and transcription (Fig. 6A). The relative distribution of categories differed among the seven-genome subclade, the remaining genomes and the full dataset. These profiles summarize gene-repertoire composition rather than transcriptional activity or formal functional enrichment.

Predicted core-protein association networks shared highly connected proteins, including GuaA, GuaB, MetG, PolA and GltB (Fig. 6B–D). Their annotations relate to nucleotide biosynthesis, translation, DNA metabolism and nitrogen metabolism. The recurrence of these proteins identifies conserved cellular functions across the groups. Network connectivity alone does not identify the molecular determinants of H380 metal tolerance.

The seven annotated subclade-associated candidates were HisM, HisJ, PotA, TerC, MarR, CaiB and PotC (Table 2). Their reported within-subclade protein identities ranged from 95.78% to 97.37%, with similarities above 97%. The annotations encompass amino acid and polyamine transport, transcriptional regulation, acyl-CoA-related metabolism and a TerC-family membrane protein. These candidates connect lineage-associated gene content to specific cellular processes that can be investigated experimentally.

3.7 Environmental homology searches do not resolve the H380 subclade

Environmental database searches returned best-hit protein identities of 38.43–69.04% across the seven candidates, with query coverage of 86–100% (Table 3). The original search summary described hits in more than 100 environmental metagenomes, including marine and compost datasets. The best-hit identities were substantially below the within-subclade conservation values. Accordingly, the searches recover broader protein-family homologues but do not establish the presence, abundance or habitat distribution of the H380 subclade. These results define a practical boundary for using the listed genes as environmental tracking markers.

4. Discussion

The principal contribution of this study is a genomic framework for interpreting a metal-tolerant soil isolate. The phenotypic overview establishes the environmental trait of interest, while the genome comparisons locate H380 within a heterogeneous *Achromobacter* lineage and separate conserved cellular functions from accessory variation. This makes the strain useful for studying how environmental tolerance is embedded within a wider genomic background.

The strong similarity between H380 and ANB-1 is more informative than their membership in a seven-genome subclade. The low ANI between these two genomes and the other subclade members shows why phylogenetic clustering must be interpreted alongside genome similarity. Earlier genus-wide work demonstrated the value of combining phylogenomics, ANI and accessory gene content in *Achromobacter* [6]. For H380, this combination identifies a close genomic comparator and prevents the broader subclade from being treated as a single species. A formal species assignment would additionally require comparison with the relevant type-strain genomes and current taxonomic records.

The tolerance phenotype gives the functional analysis a specific biological focus. Membrane transport and inorganic ion metabolism are plausible places to investigate metal handling, but their broad annotation categories encompass functions beyond metal resistance. Likewise, conservation of HisM, HisJ, PotA, PotC, MarR, CaiB and TerC within the selected subclade establishes a distribution pattern, not a causal explanation. The decisive next comparison is whether candidate genes or their regulatory responses distinguish H380 from closely related strains with measured differences in tolerance. That comparison would connect accessory variation to phenotype more directly than the connectivity of housekeeping proteins in a predicted association network.

Environmental application also requires separating growth tolerance from remediation performance. Previous work on *Achromobacter* has examined specific transformations and bioremediation-associated traits [3, 5, 32]. Here, the genome and phenotype characterize persistence-related properties; they do not quantify metal removal, immobilization or performance in soil. The accompanying antibiotic-growth observations also warrant characterization alongside any application claim. Recent clinical work shows that antimicrobial phenotypes and resistance-gene repertoires vary across *Achromobacter* species [7]; standardized measurements and gene-context analysis are therefore needed before interpreting H380 as an environmental deployment candidate.

The comparative design defines two further interpretive boundaries. First, the seven-genome and 28-genome groups differ in both size and phylogenetic breadth, so differences in core and pan-genome size cannot be attributed solely to adaptation. Assembly fragmentation can additionally influence inferred gene absence and gene-order patterns. Second, the environmental searches retrieved relatively distant homologues of broadly distributed protein functions. Such matches inform protein-family distribution but require more specific sequence markers and traceable sample metadata for lineage-level ecological inference.

Taken together, the study links a metal-tolerant environmental isolate to a close genomic relative, a defined accessory repertoire and a set of candidate transport and regulatory functions. This provides a focused starting point for testing the genetic contributions to H380 persistence under metal exposure.

5. Conclusions

Whole-genome and comparative analyses place *Achromobacter* sp. H380 close to ANB-1 within a genetically heterogeneous seven-genome subclade. H380 carries 461 strain-specific gene families in the analyzed 35-genome dataset, while selected subclade-associated genes encode transport, regulatory and metabolic functions. Together with the recorded tolerance phenotype, these findings establish a genomic context for investigating environmental persistence and prioritizing candidate functions for validation.

Abbreviations

AAI, average amino acid identity; ANI, average nucleotide identity; BRIG, BLAST Ring Image Generator; COG, Clusters of Orthologous Groups; GO, Gene Ontology; KEGG, Kyoto Encyclopedia of Genes and Genomes; NR, non-redundant protein database; OD600, optical density at 600 nm; PGAP, Prokaryotic Genome Annotation Pipeline.

Funding information: This study was supported by the Hubei Provincial Science and Technology Innovation Program Project (2024BBA001); Central Government Guidance Fund for Local Science and Technology Development (2025BBA006).

Author statement: Song Liu: investigation, conceptualization, writing-original draft, data curation. Fan Chen: investigation, methodology. Hao Gu: software. Changfan Zhou: data curation and software. Wei Zeng: data curation. Hongwei Xiao: methodology. Yanzhen Bi: supervision, resources, writing-review, and editing. Hongyan Ren: conceptualization, supervision, resources, writing-review, and editing.

Declarations

Ethics approval and consent to participate

Not applicable. This study did not involve human participants, human samples, or animals.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Data availability

The genome assembly of *Achromobacter* sp. H380 has been submitted to NCBI under BioProject accession **PRJNA1551060** and BioSample accession **SAMN64461439**. The genome submission is currently being processed, and an assembly accession number has not yet been assigned. The accession numbers of the publicly available genomes used for comparative analyses are listed in Table 1.

References

1. Bruins MR, Kapil S, Oehme FW. Microbial resistance to metals in the environment. Ecotoxicol Environ Saf. 2000;45(3):198-207. doi:10.1006/eesa.1999.1860.

2. Ayangbenro AS, Babalola OO. A New Strategy for Heavy Metal Polluted Environments: A Review of Microbial Biosorbents. Int J Environ Res Public Health. 2017;14(1):E94. doi:10.3390/ijerph14010094.

3. Cai L, Rensing C, Li X, Wang G. Novel gene clusters involved in arsenite oxidation and resistance in two arsenite oxidizers: Achromobacter sp. SY8 and Pseudomonas sp. TS44. Appl Microbiol Biotechnol. 2009;83(4):715-725. doi:10.1007/s00253-009-1929-4.

4. Deng MC, Li J, Liang FR, Yi M, Xu XM, Yuan JP, Peng J, Wu CF, Wang JH. Isolation and characterization of a novel hydrocarbon-degrading bacterium Achromobacter sp. HZ01 from the crude oil-contaminated seawater at the Daya Bay, southern China. Mar Pollut Bull. 2014;83(1):79-86. doi:10.1016/j.marpolbul.2014.04.018.

5. Hong YH, Ye CC, Zhou QZ, Wu XY, Yuan JP, Peng J, Deng H, Wang JH. Genome Sequencing Reveals the Potential of Achromobacter sp. HZ01 for Bioremediation. Front Microbiol. 2017;8:1507. doi:10.3389/fmicb.2017.01507.

6. Jeukens J, Freschi L, Vincent AT, Emond-Rheault JG, Kukavica-Ibrulj I, Charette SJ, Levesque RC. A Pan-Genomic Approach to Understand the Basis of Host Adaptation in Achromobacter. Genome Biol Evol. 2017;9(4):1030-1046. doi:10.1093/gbe/evx061.

7. Ray S, Flemming LK, Scudder CJ, Ly MA, Porterfield HS, Smith RD, Clark AE, Johnson JK, Das S. Comparative phenotypic and genotypic antimicrobial susceptibility surveillance in Achromobacter spp. through whole genome sequencing. Microbiol Spectr. 2025;13(4):e02527-24. doi:10.1128/spectrum.02527-24.

8. Huang Q, Chen W, Wang F, Huang B. Achromobacter for heavy metal cadmium pollution control and application thereof. Chinese patent CN105062926A. Published 18 November 2015. https://patents.google.com/patent/CN105062926A/en.

9. Rhoads A, Au KF. PacBio Sequencing and Its Applications. Genomics Proteomics Bioinformatics. 2015;13(5):278-289. doi:10.1016/j.gpb.2015.08.002.

10. Tatusova T, DiCuccio M, Badretdin A, Chetvernin V, Nawrocki EP, Zaslavsky L, Lomsadze A, Pruitt KD, Borodovsky M, Ostell J. NCBI prokaryotic genome annotation pipeline. Nucleic Acids Res. 2016;44(14):6614-6624. doi:10.1093/nar/gkw569.

11. Lagesen K, Hallin P, Rødland EA, Staerfeldt HH, Rognes T, Ussery DW. RNAmmer: consistent and rapid annotation of ribosomal RNA genes. Nucleic Acids Res. 2007;35(9):3100-3108. doi:10.1093/nar/gkm160.

12. Dhillon BK, Laird MR, Shay JA, Winsor GL, Lo R, Nizam F, Pereira SK, Waglechner N, McArthur AG, Langille MG, Brinkman FS. IslandViewer 3: more flexible, interactive genomic island discovery, visualization and analysis. Nucleic Acids Res. 2015;43(W1):W104-8. doi:10.1093/nar/gkv401.

13. Tatusov RL, Galperin MY, Natale DA, Koonin EV. The COG database: a tool for genome-scale analysis of protein functions and evolution. Nucleic Acids Res. 2000;28(1):33-36. doi:10.1093/nar/28.1.33.

14. Harris MA, Clark J, Ireland A, Lomax J, Ashburner M, Foulger R, Eilbeck K, Lewis S, Marshall B, Mungall C, Richter J, Rubin GM, Blake JA, Bult C, Dolan M, Drabkin H, Eppig JT, Hill DP, Ni L, Ringwald M, Balakrishnan R, Cherry JM, Christie KR, Costanzo MC, Dwight SS, Engel S, Fisk DG, Hirschman JE, Hong EL, Nash RS, Sethuraman A, Theesfeld CL, Botstein D, Dolinski K, Feierbach B, Berardini T, Mundodi S, Rhee SY, Apweiler R, Barrell D, Camon E, Dimmer E, Lee V, Chisholm R, Gaudet P, Kibbe W, Kishore R, Schwarz EM, Sternberg P, Gwinn M, Hannick L, Wortman J, Berriman M, Wood V, de la Cruz N, Tonellato P, Jaiswal P, Seigfried T, White R, Gene Ontology Consortium. The Gene Ontology (GO) database and informatics resource. Nucleic Acids Res. 2004;32(Database issue):D258-61. doi:10.1093/nar/gkh036.

15. Kanehisa M, Goto S, Sato Y, Furumichi M, Tanabe M. KEGG for integration and interpretation of large-scale molecular data sets. Nucleic Acids Res. 2012;40(Database issue):D109-14. doi:10.1093/nar/gkr988.

16. Contreras-Moreira B, Vinuesa P. GET_HOMOLOGUES, a versatile software package for scalable and robust microbial pangenome analysis. Appl Environ Microbiol. 2013;79(24):7696-7701. doi:10.1128/aem.02411-13.

17. Li L, Stoeckert CJ, Roos DS. OrthoMCL: identification of ortholog groups for eukaryotic genomes. Genome Res. 2003;13(9):2178-2189. doi:10.1101/gr.1224503.

18. Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 2004;32(5):1792-1797. doi:10.1093/nar/gkh340.

19. Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. Bioinformatics. 2009;25(15):1972-1973. doi:10.1093/bioinformatics/btp348.

20. Nguyen LT, Schmidt HA, von Haeseler A, Minh BQ. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol Biol Evol. 2015;32(1):268-274. doi:10.1093/molbev/msu300.

21. Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, Jermiin LS. ModelFinder: fast model selection for accurate phylogenetic estimates. Nat Methods. 2017;14(6):587-589. doi:10.1038/nmeth.4285.

22. Bryant D, Moulton V. Neighbor-net: an agglomerative method for the construction of phylogenetic networks. Mol Biol Evol. 2004;21(2):255-265. doi:10.1093/molbev/msh018.

23. Huson DH, Bryant D. Application of phylogenetic networks in evolutionary studies. Mol Biol Evol. 2006;23(2):254-267. doi:10.1093/molbev/msj030.

24. Yoon SH, Ha SM, Lim J, Kwon S, Chun J. A large-scale evaluation of algorithms to calculate average nucleotide identity. Antonie Van Leeuwenhoek. 2017;110(10):1281-1286. doi:10.1007/s10482-017-0844-4.

25. Alikhan NF, Petty NK, Ben Zakour NL, Beatson SA. BLAST Ring Image Generator (BRIG): simple prokaryote genome comparisons. BMC Genomics. 2011;12:402. doi:10.1186/1471-2164-12-402.

26. Wang Y, Tang H, Debarry JD, Tan X, Li J, Wang X, Lee TH, Jin H, Marler B, Guo H, Kissinger JC, Paterson AH. MCScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. Nucleic Acids Res. 2012;40(7):e49. doi:10.1093/nar/gkr1293.

27. Tettelin H, Masignani V, Cieslewicz MJ, Donati C, Medini D, Ward NL, Angiuoli SV, Crabtree J, Jones AL, Durkin AS, Deboy RT, Davidsen TM, Mora M, Scarselli M, Margarit y Ros I, Peterson JD, Hauser CR, Sundaram JP, Nelson WC, Madupu R, Brinkac LM, Dodson RJ, Rosovitz MJ, Sullivan SA, Daugherty SC, Haft DH, Selengut J, Gwinn ML, Zhou L, Zafar N, Khouri H, Radune D, Dimitrov G, Watkins K, O'Connor KJ, Smith S, Utterback TR, White O, Rubens CE, Grandi G, Madoff LC, Kasper DL, Telford JL, Wessels MR, Rappuoli R, Fraser CM. Genome analysis of multiple pathogenic isolates of Streptococcus agalactiae: implications for the microbial "pan-genome". Proc Natl Acad Sci U S A. 2005;102(39):13950-13955. doi:10.1073/pnas.0506758102.

28. Vernikos G, Medini D, Riley DR, Tettelin H. Ten years of pan-genome analyses. Curr Opin Microbiol. 2015;23:148-154. doi:10.1016/j.mib.2014.11.016.

29. Wu S, Zhu Z, Fu L, Niu B, Li W. WebMGA: a customizable web server for fast metagenomic sequence analysis. BMC Genomics. 2011;12:444. doi:10.1186/1471-2164-12-444.

30. Szklarczyk D, Morris JH, Cook H, Kuhn M, Wyder S, Simonovic M, Santos A, Doncheva NT, Roth A, Bork P, Jensen LJ, von Mering C. The STRING database in 2017: quality-controlled protein-protein association networks, made broadly accessible. Nucleic Acids Res. 2017;45(D1):D362-D368. doi:10.1093/nar/gkw937.

31. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res. 2003;13(11):2498-2504. doi:10.1101/gr.1239303.

32. Mohapatra B, Kazy SK, Sar P. Comparative genome analysis of arsenic reducing, hydrocarbon metabolizing groundwater bacterium Achromobacter sp. KAs 3-5T explains its competitive edge for survival in aquifer environment. Genomics. 2019;111(6):1604-1619. doi:10.1016/j.ygeno.2018.11.004.

Tables

Table 1 Genome assemblies included in the comparison

Strain	Recorded species	Assembly accession	Assembly level
NCTC10807	A. xylosoxidans	GCA_001457475.1	Complete
NBRC 15126	A. xylosoxidans	GCA_000508285.1	Complete
A8	A. xylosoxidans	GCA_000165835.1	Complete
AX1	A. xylosoxidans	GCA_008432465.1	Complete
MN001	A. xylosoxidans	GCA_001051055.1	Complete
NH44784-1996	A. xylosoxidans	GCA_000967095.1	Chromosome
FDAARGOS_147	A. xylosoxidans	GCA_001558755.2	Complete
DN002	A. xylosoxidans	GCA_009363015.1	Complete
SOLR10	A. xylosoxidans	GCA_003293615.1	Chromosome
AXX-A	A. insuavis	GCA_000219745.1	Scaffold
ANB-1	A. pulmonis	GCA_002885955.2	Scaffold
DSM 23807	A. insolitus	GCA_001971645.1	Complete
NCTC 13520	A. insolitus	GCA_900637265.1	Complete
FDAARGOS_88	A. insolitus	GCA_000783435.2	Complete
LCu2	A. insolitus	GCA_008245125.1	Complete
AB2	A. insolitus	GCA_003293535.1	Chromosome
PheN1	A. denitrificans	GCA_003812265.1	Complete
PR1	A. denitrificans	GCA_002205315.1	Complete
USDA-ARS-USMARC-56712	A. denitrificans	GCA_001514355.1	Complete
AD_173	A. denitrificans	GCA_003939895.1	Scaffold
MT3	A. denitrificans	GCA_001945385.1	Scaffold
LMG 3411	A. agilis	GCA_900496965.1	Contig
KCJK 8006	A. mucicolens	GCA_003053425.1	Contig
DSM 23806	A. spanius	GCA_002812705.1	Complete
UQ283	A. spanius	GCA_003994415.1	Complete
NCTC13519	A. spanius	GCA_900636675.1	Complete
MYb73	A. spanius	GCA_002966795.1	Complete
NBRC102461	A. piechaudii	GCA_001571245.1	Contig
GCS2	A. piechaudii	GCA_001262515.1	Scaffold
ATCC43553	A. piechaudii	GCA_000164035.1	Scaffold
270	A. aegrifaciens	GCA_003939885.1	Scaffold
B2	A. marplatensis	GCA_002209535.1	Contig
CECT7342	A. marplatensis	GCA_003315095.1	Scaffold
HLE	A. marplatensis	GCA_000286415.1	Contig
H380	*Achromobacter* sp.	[D01 pending]	Circular chromosome reported

Species labels and assembly levels refer to the original dataset and have not been reclassified. H380 was newly sequenced for this study; its public accession must be supplied.

Table 2 Annotated candidates associated with the seven genome subclade

Protein	Predicted function	Length (aa)	Identity (%)	Similarity (%)
HisM	Amino acid ABC transporter permease	246	95.98	97.568
HisJ	Amino acid transport substrate-binding component	267	97.35	98.319
PotA	Polyamine ABC transporter ATPase	286	96.62	97.42
TerC	TerC-family membrane protein	369	96.09	97.45
MarR	Transcriptional regulator	177	97.37	97.37
CaiB	Predicted acyl-CoA transferase/carnitine-related protein	421	95.78	97.18
PotC	Polyamine ABC transporter permease	233	95.81	97.61

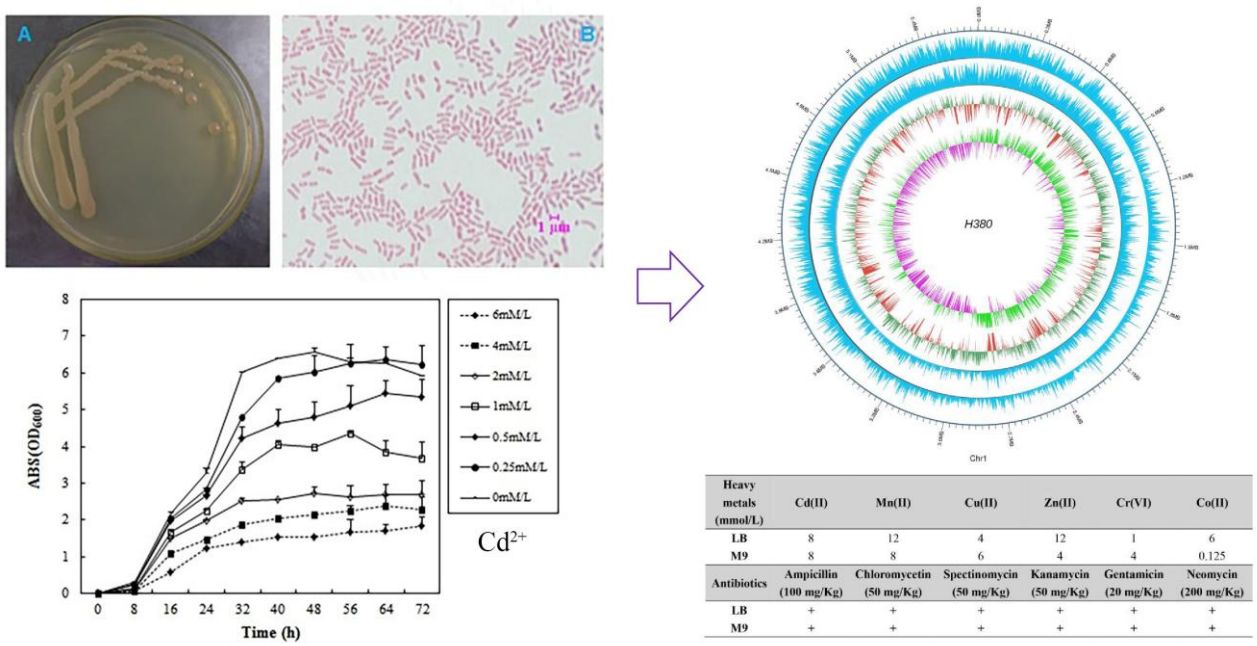
Identity and similarity are the reported within-subclade protein values. The scope of averaging and the corresponding locus tags require confirmation from the original alignments (M09). Functional labels are annotations, not experimental assignments.

Table 3 Reported best environmental protein matches

Query protein	Query coverage (%)	Protein identity (%)
HisM	98	69.04
HisJ	96	45.95
PotA	95	38.43
TerC	86	52.85
MarR	91	44.38
CaiB	100	50
PotC	98	55.31

Values reproduce the original env_nr search summary. They describe the best reported match for each query, not within-subclade conservation or environmental abundance. Hit accessions and search settings are requested in M09.

Figure 1 Phenotypic overview and genome map of H380

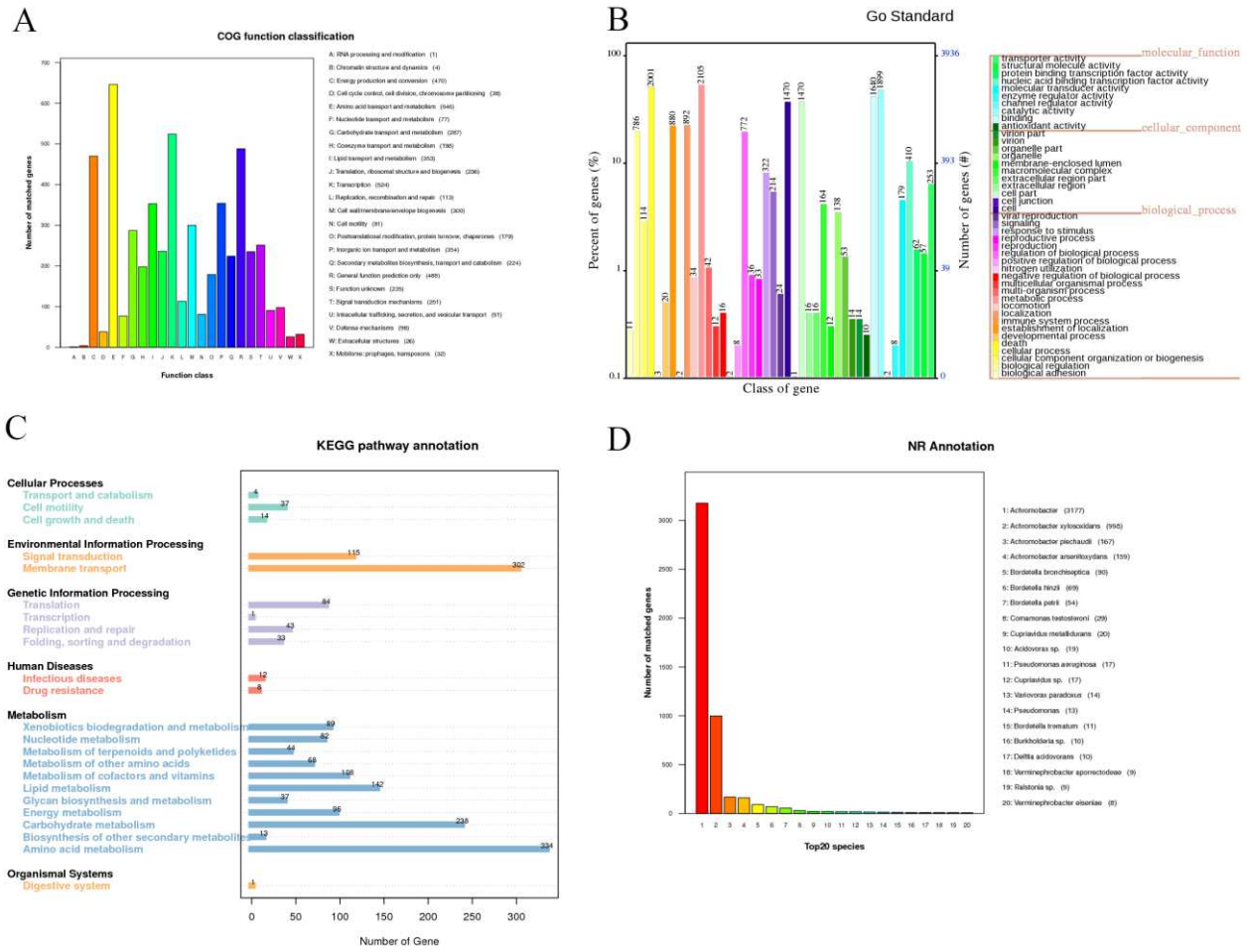


Heavy metals (mmol/L)	Cd(II)	Mn(II)	Cu(II)	Zn(II)	Cr(VI)	Co(II)
LB	8	12	4	12	1	6
M9	8	8	6	4	4	0.125

Antibiotics	Ampicillin (100 mg/Kg)	Chloromycetin (50 mg/Kg)	Spectinomycin (50 mg/Kg)	Kanamycin (50 mg/Kg)	Gentamicin (20 mg/Kg)	Neomycin (200 mg/Kg)
LB	+	+	+	+	+	+
M9	+	+	+	+	+	+

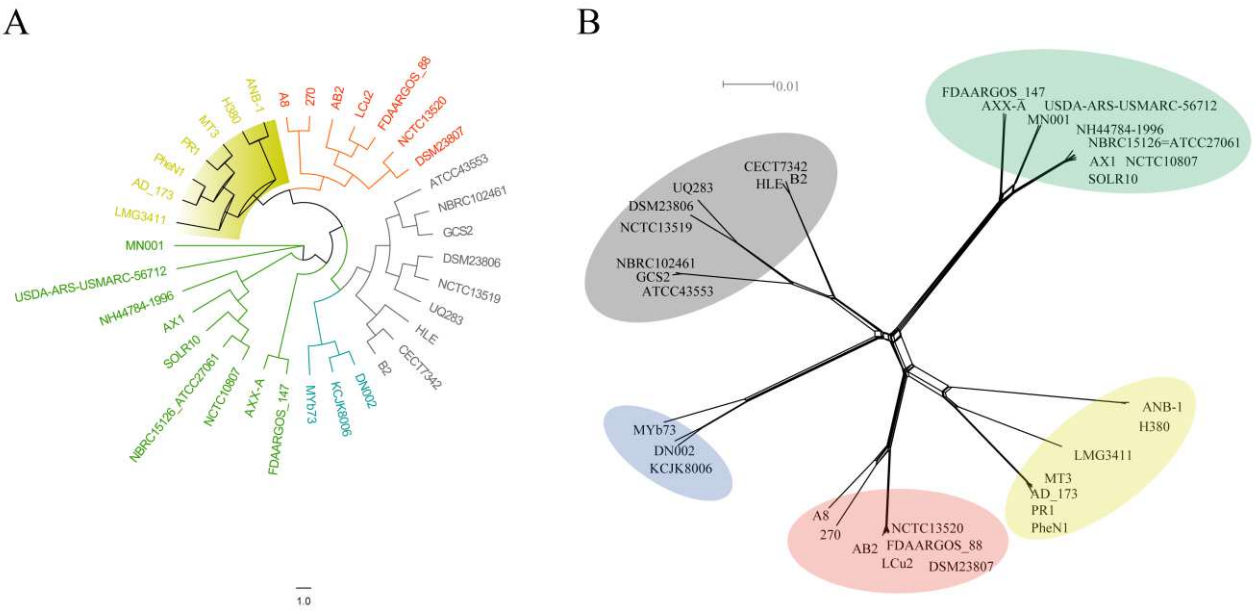
Upper left, colony morphology (A) and microscopy (B; displayed scale bar, 1 μm). Lower left, OD600 time courses during cadmium exposure over 72 h. Upper right, circular chromosome map. Lower right, the recorded metal-tolerance values and qualitative growth at the displayed antibiotic concentrations in LB and M9. The image is reproduced intact from the original graphical overview.

Figure 2 Functional annotation of the H380 genome



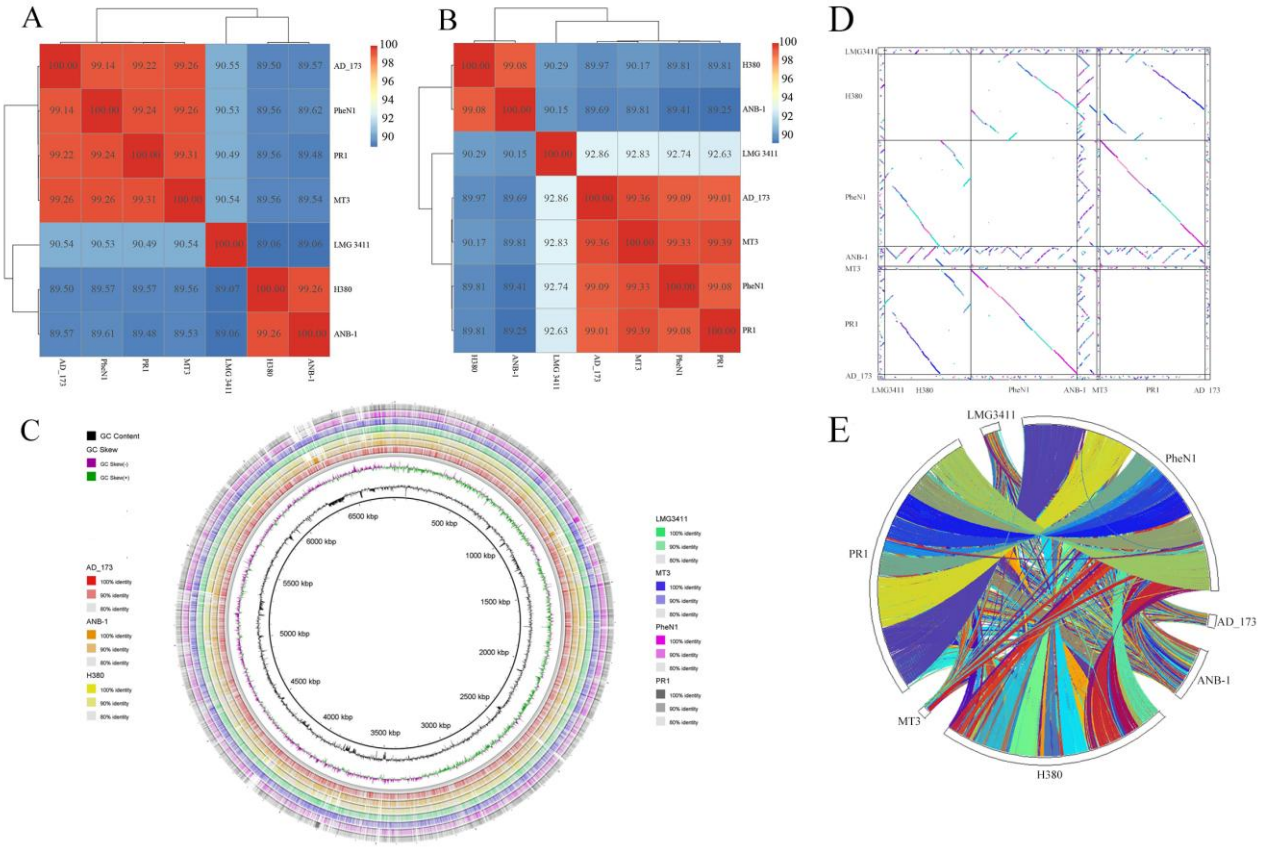
(A) COG functional categories. (B) GO categories. (C) KEGG pathway assignments. (D) NR protein annotation summary. Counts represent annotation assignments; they do not measure gene expression or pathway activity.

Figure 3 Phylogenetic relationships among the 35 Achromobacter genomes



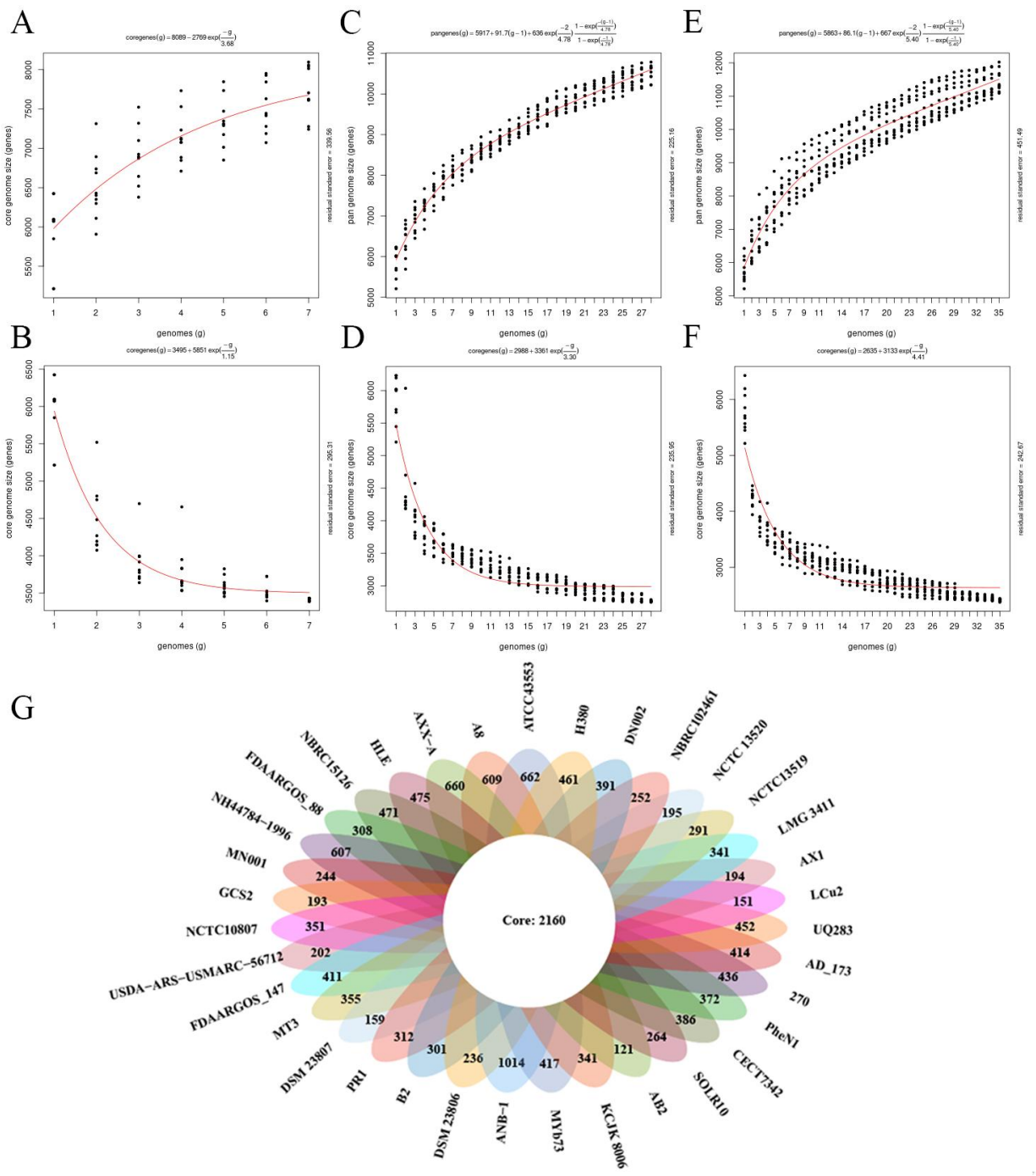
(A) Maximum-likelihood tree inferred from the shared single-copy orthologue dataset. (B) Neighbor-Net network. The H380-containing seven-genome subclade is highlighted in yellow. Other colors identify groups displayed in the original analysis.

Figure 4 Genome similarity and gene-order comparisons within the H380 subclade



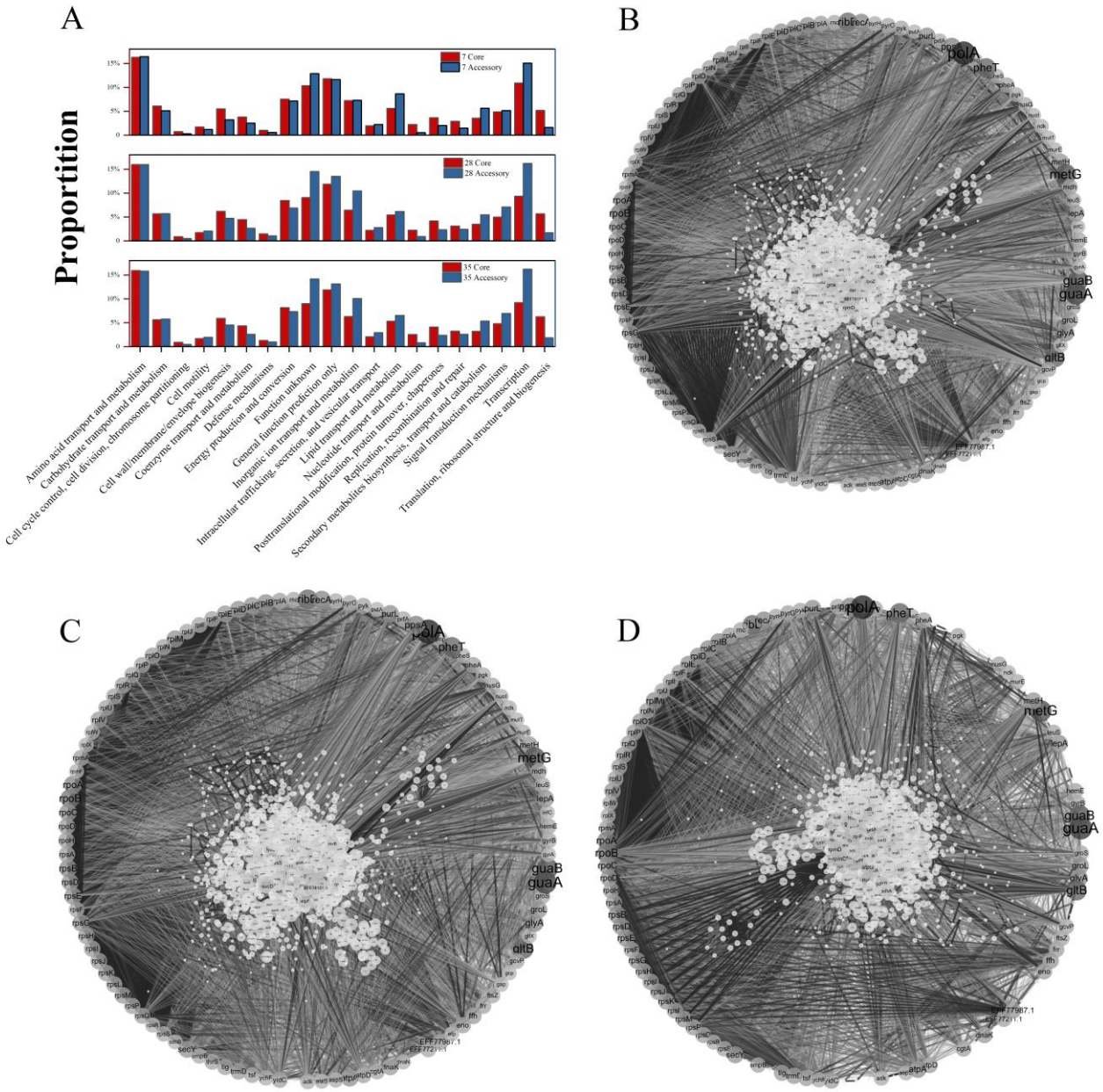
(A) ANI and (B) AAI heatmaps for seven genomes. Values are percentages. (C) BRIG comparison with the reference genome; shading denotes the displayed identity bands. (D) Gene-order dot plot. (E) Circular representation of collinear gene blocks. The dataset includes fragmented assemblies.

Figure 5 Core and pan-genome repertoires at three sampling scales



Pan-genome accumulation and core-genome curves for (A,B) the H380 subclade (n = 7), (C,D) the remaining genomes (n = 28), and (E,F) all genomes (n = 35). (G) Shared-core and strain-specific family counts shown in the original graphic.

Figure 6 Functional composition and predicted core-protein associations



(A) COG category proportions for core and accessory gene sets in the H380 subclade, the remaining genomes and the complete dataset. Predicted core-protein association networks for (B) the H380 subclade, (C) the remaining genomes and (D) all genomes. Networks were generated using STRING and displayed in Cytoscape.