bioRxiv preprint doi: https://doi.org/10.1101/2021.08.29.457868; this version posted August 30, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Biosynthetic gene cluster profiling predicts the positive association between antagonism and phylogeny in *Bacillus*

Authors:

Liming Xia[1§], Youzhi Miao[1§], A'li Cao[1], Yan Liu[1], Zihao Liu[1], Xinli Sun[1], Yansheng Xue[1], Zhihui Xu[1], Weibing Xun[1], Qirong Shen[1], Nan Zhang[1*], Ruifu Zhang[2*]

Affiliations:

[1]Jiangsu Provincial Key Lab of Solid Organic Waste Utilization, Jiangsu Collaborative Innovation Center of Solid Organic Wastes, Educational Ministry Engineering Center of Resource-saving fertilizers, Nanjing Agricultural University, Nanjing 210095, Jiangsu, P. R. China

[2]Key Laboratory of Microbial Resources Collection and Preservation, Ministry of Agriculture, Institute of Agricultural Resources and Regional Planning, Chinese Academy of Agricultural Sciences, Beijing 100081, P. R. China

[§]Both authors contributed equally to this paper.

[*]**Corresponding authors:**

Nan Zhang, College of Resources & Environmental Sciences, Nanjing Agricultural University, 210095, Nanjing, P. R. China E-mail: nanzhang@njau.edu.cn.

Ruifu Zhang, Institute of Agricultural Resources and Regional Planning, Chinese Academy of Agricultural Sciences, Beijing 100081, P. R. China; E-mail: zhangruifu@caas.cn.

The authors declare no conflict of interest.

bioRxiv preprint doi: https://doi.org/10.1101/2021.08.29.457868; this version posted August 30, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Abstract: Understanding the driving forces and intrinsic mechanisms of microbial competition is a fundamental question in microbial ecology. Despite the well-established negative correlation between exploitation competition and phylogenetic distance, the process of interference competition that is exemplified by antagonism remains controversial. Here, we studied the genus *Bacillus*, a commonly recognized producer of multifarious antibiotics, to explore the role of phylogenetic patterns of biosynthetic gene clusters (BGCs) in mediating the relationship between antagonism and phylogeny. Comparative genomic analysis revealed a positive association between BGC distance and phylogenetic distance. Antagonistic tests demonstrated that the inhibition phenotype positively correlated with both phylogenetic and predicted BGC distance, especially for antagonistic strains possessing abundant BGCs. Mutant-based verification showed that the antagonism was dependent on the BGCs that specifically harbored by the antagonistic strain. These findings highlight that BGC-phylogeny coherence regulates the positive correlation between congeneric antagonism and phylogenetic distance, which deepens our understanding of the driving force and intrinsic mechanism of microbial interactions.

Keywords: antagonism; biosynthetic gene clusters (BGCs); phylogenetic distance; *Bacillus*

Introduction

Microbes are naturally surrounded by taxonomically different ones with which they compete for scarce resources and space[1]. Competition between different species is generally categorized as exploitation competition, which involves the rapid consumption of a limited resource[2,3], and interference competition, which refers to direct antagonistic interactions[4]. Microbial antagonism is

bioRxiv preprint doi: https://doi.org/10.1101/2021.08.29.457868; this version posted August 30, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

41 driven by diverse toxins, such as broad-spectrum antibiotics and strain-specific bacteriocins[5,6], and
42 is recognized as a key element regulating populations and determining their success within diverse
43 communities[7]. The stunning diversity of both the categories and functions of antimicrobial
44 metabolites among different species, results in the extraordinary complexity of interference
45 competition between microbes with different phylogenetic relationships[4,8]. Accordingly, illustrating
46 the driving forces and mechanisms of antagonistic competition is crucial for understanding and
47 predicting microbial behaviors during community assemblage and succession[2].

48 Phylogenetic relatedness is considered to be closely associated with microbial competition,
49 based on its determination of both primary and secondary metabolic profiles[2,4]. Despite the well-
50 established negative correlation between exploitation competition and phylogenetic distance[2], the
51 process of interference competition is much more complicated and remains controversial. Russel et
52 al. demonstrated that inhibition was more prevalent between closely related bacteria, and this
53 negative correlation between antagonism and phylogeny was mediated by the overlap of the
54 metabolic niche among different strains[9]. Conversely, other studies examined congeneric
55 competition in *Vibrio* and *Streptomyces* and revealed that closely related strains competed less than
56 phylogenetically distant strains, which was probably caused by the effect of the prior coexistence
57 and distribution of secondary metabolites in different genomes[7,10,11]. Additionally, the positive
58 relationship between kin discrimination and phylogeny was indicated within *Bacillus subtilis*, which
59 was modulated by genes involved in antimicrobials and cell-surface modifiers[12,13]; however, this
60 correlation was lost or even became a certain extent of negative when more distantly *Bacillus* strains
61 were tested for antagonism, probably being dependent on the demand of protecting public goods[14].

bioRxiv preprint doi: https://doi.org/10.1101/2021.08.29.457868; this version posted August 30, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Taken together, the relationship between antagonism and phylogenetic distance with regard to microbes from different taxonomical scale or groups, as well as the involved biological mechanism, is still under debate, which limits both our understanding and application of these microbial interactions.

Biosynthetic gene clusters (BGCs) are responsible for the production of various secondary metabolites that contribute to interference competition between different microbes[15,16], and also usually provide resistance against the self-produced antibiotic to protect the host cell[4,17,18]. Although the relevance (or lack of relevance) of BGCs to antagonism or phylogeny have been evaluated in diverse microbes[11,16,17,19-21], the involvement of BGCs profile in mediating the relationship between interference competition and phylogeny, has not been well addressed. Here, we hypothesize that the correlation between BGC and phylogenetic distance can predict the pattern of congeneric antagonism among different taxonomic groups, as strains possessing higher BGC similarity should have a lower probability of inhibiting each other. To test this hypothesis, we referred to the Gram-positive *Bacillus* as the target genus, which is a commonly recognized producer of multifarious secondary metabolites[16,19,22-25], including non-ribosomal lipopeptides (e.g., surfactin, iturin, and fengycin families produced by various species)[23], non-ribosomal polyketides (e.g., difficidin and macrolactin produced by *B. amyloliquefaciens* and *B. velezensis*)[19], peptide-polyketide hybrid compound (e.g., zwittermicin produced by *B. cereus* and *B. thuringiensis*)[24], and ribosomally synthesized and post-translationally modified peptides (RiPPs; e.g., lichenicidin produced by *B. licheniformis*)[26]. We attempted to address the role of the BGC-phylogeny relevance in shaping the association between antagonism and phylogenetic distance based on comparative genomic analysis,

bioRxiv preprint doi: https://doi.org/10.1101/2021.08.29.457868; this version posted August 30, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

antagonistic assessments, and mutant-based verification. This study provides new insights to better understand the driving force and mechanism of interference competition in microbial ecology, which will allow us to better manipulate community assemblages for practical purpose.

Results

Positive correlation between biosynthetic gene cluster (BGC) and phylogenetic distance in the genus *Bacillus*

BGCs are responsible for the synthesis of secondary metabolites involved in microbial interference competition. To investigate the relationship between BGC and phylogenetic distance within the genus *Bacillus*, we collected 4,268 available *Bacillus* genomes covering 139 species from the NCBI database (Table S1). Phylogenetic analysis based on the sequences of 120 ubiquitous single-copy proteins[27] showed that the 139 species could be generally clustered into four clades (Fig. 1 & Table S2; the phylogenetic tree including all the detailed species information is shown in Fig. S1), including a *subtilis* clade that includes species from diverse niches and can be further divided into the *subtilis* and *pumilus* subclades, a *cereus* clade that contains typical pathogenic species (*B. cereus*, *B. anthracis*, *B. thuringiensis*, etc.), a *megaterium* clade, and a *circulans* clade.

Prediction using the bioinformatic tool antiSMASH[15] detected 49,671 putative BGCs in the 4,268 genomes, corresponding to an average of 11.6 BGCs per genome (Table S3). The *subtilis* clade had the most BGCs, 13.1 BGCs per genome, followed by the *cereus* and *megaterium* clades, while the *circulans* clade only had 4.3 BGCs/genome (Fig. 2a; Table S4). The two most abundant BGC classes were nonribosomal peptide-synthetase (NRPS) and RiPPs, which had an abundance of 3.7

bioRxiv preprint doi: https://doi.org/10.1101/2021.08.29.457868; this version posted August 30, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

and 3.1 per genome on average, respectively (Fig. S2 & Table S4). Interestingly, *subtilis* clade accommodated significantly higher abundance of BGCs in other polyketide synthase (PKSother; 2.0 per genome vs. 0.0~1.1 per genome) and PKS-NRPS_Hybrids (0.7 vs. 0.0~0.2) classes, as compared with the three other clades (Table S4); while *cereus* clade had more BGCs in RiPPs than other clades on average (Table S4). Overall, the profile of BGC products and classification was generally consistent with the phylogenetic tree (Fig. S3).

To further evaluate whether the diversity and concrete distribution of the BGCs among genomes were relevant to the phylogenetic relatedness, we selected 545 representative *Bacillus* genomes based on the following criteria: (i) high genome sequencing quality for further BGC distance calculation, and (ii) covering all *Bacillus* species. We analyzed the interactive sequence similarity network of BGCs in these genomes by using the "biosynthetic gene similarity clustering and prospecting engine" (BiG-SCAPE)[28]. The (dis)similarity of paired BGCs was calculated based on a combination of three metrics, including the Jaccard index (JI), adjacency index (AI), and domain sequence similarity (DSS), which resulted in 1,110 gene cluster families (GCFs) and 76 gene cluster clans (GCCs) of the 4,877 putative BGCs (Tables S5 & S6). The hierarchal clustering based on the abundance of these GCFs among each genome (Table S7) indicated that, each phylogenetic clade/subclade revealed its own distinctive BGCs distribution profile, and possessed a number of taxonomy-specific secondary metabolites (Figs. 2b & S4; Table S5).The widespread BGCs in *cereus* clade included fengycin, bacillibactin, bacteriocin, NRPS, and petrobactin, in which petrobactin was a clade-specific BGC; polyoxypeptin, thurincin, and zwittermicin were also specific molecules but were mainly present in a certain of *B. cereus* and *B. thuringiensis* genomes (Figs. 2b & S4). In *subtilis*

bioRxiv preprint doi: https://doi.org/10.1101/2021.08.29.457868; this version posted August 30, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

clade, most species possessed surfactin, fengycin, bacilysin, bacillibactin, and T3PKS, while each group can also produce unique BGCs, such as betalactone for *B. pumilus*, subtilin and subtilosin for *B. subtilis*, difficidin and macrolactin for *B. amyloliquefaciens* and *B. velezensis*, mersacidin, plantazolicin, and plipastatin for a certain of genomes in the above two species, and lichenysin for *B. licheniformis* (Figs. 2b & S4). Species in *megaterium* clade mostly accommodated siderophore, surfactin, and T3PKS, some strains can potentially produce lanthipeptide, paeninodin, or bacteriocins. The dominating BGC in *circulans* clade was identified as T3PKS, and some species may synthesize siderophore, bacteriocin, or lanthipeptide (Figs. 2b & S4).

Furthermore, we calculated the BGC distance between different genomes based on the above GCFs clustering data, and found a significant positive correlation between the BGC and phylogenetic distance (Fig. 2c) ($P < 0.001$, $R^2 = 0.2847$). Interestingly, the BGC distance of genomes within or between closely related species (phylogenetic distance < 0.1) was all very close (Figs. 2c & S5); while the distance for distant species (phylogenetic distance > 0.3) became much dispersive (Figs. 2c & S5), as some can be relatively close (e.g., a certain of connections between *cereus* and *subtilis* clade, Fig. S5d) but some can be very remote (e.g., a certain of connections between *circulans* clade and other three clades, Fig. S5h). To summarize, these findings demonstrate that the BGCs distribution profile was generally dependent on the phylogenetic relationship within the genus *Bacillus*.

Antagonism positively correlates with both the phylogenetic and BGC distance in *Bacillus*

BGCs not only contribute to the synthesis of secondary metabolites but also usually afford self-tolerance to the antibiotic they encode[29]. We therefore hypothesized that the BGC-phylogeny

bioRxiv preprint doi: https://doi.org/10.1101/2021.08.29.457868; this version posted August 30, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

146 coherence in *Bacillus* (Fig. 2b & 2c) determines a positive correlation between antagonism and
147 phylogenetic distance. To verify this hypothesis, we first used the bacterial colony confrontation
148 assay to investigate the relationship between the antagonistic efficiency and phylogenetic distance
149 of the paired strains (Fig. S6). The antagonistic bacteria included eight strains from the *subtilis* or
150 *cereus* clade, which are the two dominant groups within the genus *Bacillus* and have been explored
151 as providing abundant secondary metabolites (Fig. 2a); the target bacteria included 61 strains
152 representing all four clades with different phylogenetic similarities (Table S8). The results indicated
153 that all of these antagonistic strains tended to show stronger antagonistic ability towards distant
154 species than towards closely related species. For instance, *B. amyloliquefaciens* ACCC19745 and *B.*
155 *pumilus* ACCC04450 (both belong to the *subtilis* clade) showed weak antagonism against strains in
156 the same subclade but exhibited an increased antagonistic ability to the out-clade species (Fig. 3a &
157 3b). Correspondingly, the antagonistic abilities of *B. thuringiensis* YX7 and *B. mobilis* XL40 (*cereus*
158 clade) towards other *Bacillus* strains were also enhanced with increasing phylogenetic distance (Fig.
159 3c & 3d). Based on the results of the colony confrontation assays, a significant positive correlation
160 between antagonism and phylogenetic distance was revealed (Fig. 3e) ($P < 0.001$, $R^2 = 0.1263$). To
161 further clarify whether this association was mediated by the BGC distribution pattern, we calculated
162 the predicted BGC distance among all the tested paired strains (for strains whose genomes have not
163 been completely sequenced, we referred to the *Bacillus* genomes in the NCBI database that shared
164 the highest 16S rRNA similarity). Interestingly, there was also a significant positive correlation
165 between antagonism and the predicted BGC distance (Fig. 3f) ($P < 0.001$, $R^2 = 0.1132$), suggesting
166 that BGC profiling is likely to play a role in regulating interspecies antagonism.

bioRxiv preprint doi: https://doi.org/10.1101/2021.08.29.457868; this version posted August 30, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Furthermore, to check the positive antagonism-phylogeny correlation in a more defined system, we performed a fermentation supernatant assessment to test the inhibition between paired strains. This strategy can avoid potential bacterial nutrient competition and is feasible for a wider range of antagonistic strains since antagonism is not influenced by the growth speed or morphology of the colony (Fig. S6). Here, we expanded the antagonistic strains to 17 species covering all four phylogenetic clades (Table S8). The extracellular metabolites of antagonistic bacteria also generally showed stronger inhibition to distantly related strains than to closely related strains; this pattern was particularly clear for antagonistic strains in the *subtilis* clade, which harbored the most BGCs (Figs. 4a & S6). As expected, antagonism showed a significant positive correlation with both phylogenetic (Fig. 4b, $P < 0.001$, $R^2 = 0.1619$) and predicted BGC distance (Fig. 4c, $P < 0.001$, $R^2 = 0.0852$). Intriguingly, we found that the correlation between antagonism and phylogenetic distance for each individual antagonistic strain, was positively associated with the predicted quantity of BGCs in this bacteria (for strains whose genomes have not been completely sequenced, we referred to the average quantity of BGCs in this species; Fig. 4d, $P < 0.01$, $R^2 = 0.3663$). This finding suggests that the antagonistic bacteria carrying abundant BGCs (e.g., > 10) tend to have a clear positive correlation between inhibition phenotype and phylogenetic distance, while those with fewer BGCs (e.g., < 8) show a weak or even irregular antagonistic pattern against diverse targets. Furthermore, the BGC-phylogeny coherence was similar among all the antagonistic strains (no significant relevance with BGCs No.; Fig. S7, $P = 0.4201$, $R^2 = -0.0199$), while the antagonism-BGC distance correlation revealed a positive association with the quantity of BGCs in bacteria (i.e., bacteria having fewer BGCs showed a weak antagonism-BGC distance relevance, and vice versa.; Fig. 4e, $P < 0.05$, $R^2 = 0.1822$), which

bioRxiv preprint doi: https://doi.org/10.1101/2021.08.29.457868; this version posted August 30, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

188 can partially explain the weak correlation between antagonism and phylogenetic distance in strains
189 possessing fewer BGCs.
190
191 **The positive correlation of antagonism and phylogenetic distance in *Bacillus* is mediated by the**
192 **specifically harbored BGCs in antagonistic strains**
193 Having determined that the positive correlation between antagonism and phylogenetic distance
194 was consistent with the BGC-phylogeny coherence in *Bacillus*, we further investigated the mediation
195 mechanism of BGCs in the interspecies interactions. We used a typical secondary metabolite
196 producer, *B. velezensis* SQR9 belonging to *subtilis* clade, to identify the primary antagonistic
197 antibiotic towards different strains. Strain SQR9 devotes approximately 9.9% of its genome to the
198 synthesis of various antimicrobial metabolites[30], including five nonribosomal lipopeptides or
199 dipeptides (surfactin, bacillomycin D, fengycin, bacillibactin, and bacilysin), three polyketides
200 (macrolactin, bacillaene, and difficidin)[31], and one antimicrobial fatty acid (FA; bacillunoic acid)[18].
201 The antagonistic characteristics of SQR9 mutants deficient in each of the above BGCs and SQR9Δ*sfp*
202 with the 4'-phosphopantetheinyl transferase gene deleted and only bacilysin can be synthesized[22,30]
203 (Table S8), towards 24 target strains (Table S8) were investigated using a fermentation supernatant
204 inhibition assay. SQR9Δ*sfp* nearly completely lost its antagonism towards all the target strains,
205 suggesting that the synthesis of the antibiotics involved in congeneric antagonism is strongly
206 dependent on Sfp (Fig. 5). The active antimicrobial metabolites were found to be relevant to the
207 phylogenetic positions of the target strains, as a specific BGC was involved in the inhibition of strains
208 in one taxonomic group. In detail, difficidin dominated the suppression of the *megaterium* clade (Fig.

bioRxiv preprint doi: https://doi.org/10.1101/2021.08.29.457868; this version posted August 30, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

5); macrolactin was the primary antibiotic against the *cereus* clade (Fig. 5); difficidin and bacillaene both contributed to the inhibition of the *circulans* clade (Fig. 5). Strain SQR9 also used the antimicrobial FA to compete with strains in closely related species (*B. halotolerans* CF7, *B. licheniformis* LY2, and *B. sonorensis* YX13) (Fig. 5). Furthermore, we assigned the BGCs present in the testing strains if more than 80% of the corresponding *Bacillus* species genomes possessed the cluster (marked by the cross in Fig. 5). Importantly, antagonism was fully attributed to the BGCs that were present in strain SQR9 but absent in the target strains, while the metabolites shared by both SQR9 and the target strain were not involved. Additionally, the three identified functional antibiotics (difficidin, macrolactin, and bacillaene) for congeneric inhibition belong to PKSother or PKS-NRPS_Hybrids classes; enrichment of both classes in *subtilis* clade than the other three (Table S4) coincides with the weak inhibition on *subtilis* clade by strain SQR9. Taken together, these results demonstrated that interference competition was dependent on the BGCs specifically harbored by the antagonistic strains, and strains that shared more (analogous) BGCs tended to have a lower probability and intensity of antagonism. It should be noted that not all the unique BGCs in antagonistic strains contribute to inhibition of the targets (Fig. 5), and whether a target strain being resistant or sensitive to a specific secondary metabolite that it doesn't synthesis, shall involve complicated molecular mechanisms and still needs further investigation.

Discussion

The competition-relatedness hypothesis proposed by Charles R. Darwin in the *Origin of Species*, that is, congeneric species are likely to compete more fiercely by means of their functional similarity[32],

bioRxiv preprint doi: https://doi.org/10.1101/2021.08.29.457868; this version posted August 30, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

has been examined in various organisms and has received both positive and negative support[33-36]. With regard to the microbial fierce competition exemplified by antagonism, Russel et al. found a negative correlation between inhibition probability and phylogenetic distance[9]; some other scientists discovered a positive relationship between antagonistic interaction (including kin discrimination) and phylogenetic dissimilarity in genus *Vibrio*[7] and *Streptomyces*[10,11], and species *B. subtilis*[12,13]. The present study demonstrated that antagonism tended to be positively correlated with phylogenetic distance within the genus *Bacillus* (Figs. 3 & 4). Comprehensively, we pronounce that within a relatively narrow phylogenetic range, such as for congeneric strains, BGC similarity, which determines the secondary metabolite profile, appears to be the main driver of antagonistic interactions (Figs. 2~5). BGCs themselves, or other elements in the same genome, usually afford self-resistance by providing active efflux or modification of the relevant antibiotic[29,37], while the absence of a specific BGC suggests the potential to be sensitive to this metabolite[17,18]; this principle was also confirmed by the observation that antagonism was dependent on the BGC that was present in the antagonistic strain but absent in the target strain (Fig. 5). In addition, despite the mobility of BGCs among different microbes[16,38], their distribution pattern was generally in accordance with the phylogenetic relationship within genus (Fig. 2b & 2c). Consequently, closely related species with a higher BGC similarity have a lower probability of inhibiting each other, while distant species in the same genus are likely to suppress each other more fiercely (Figs. 3 & 4). Therefore, we highlighted that the coherence between BGCs distribution and phylogenetic characteristics is one of the crucial factors regulating congeneric interactions. Comparatively, at a larger taxonomic scale, the significant variation in the BGCs distribution among bacteria from different genera or even phyla

bioRxiv preprint doi: https://doi.org/10.1101/2021.08.29.457868; this version posted August 30, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

can contribute to an irregular correlation between the secondary metabolite profile and phylogenetic distance[28,39,40]. In this situation, functional similarity, such as metabolic niche overlap, may become the main driver and lead to a negative correlation between antagonism and phylogenetic distance[9]. Taken together, this taxonomic range-dependent association between antagonism and phylogeny should indicate a cooperation-competition tradeoff during microbial interactions and is coordinated by a set of sophisticated molecular mechanisms.

Interestingly, the positive correlation between antagonism and phylogeny was relatively strong in antagonistic strains possessing abundant BGCs (e.g., > 10), but was weak or even not significant in those harboring fewer BGCs (e.g., < 8) (Fig. 4d). This divergence is likely to arise from the altered antagonism-BGC distance correlation in different bacteria (Figs. 4e & S6). As inhibition is mainly dependent on specific BGCs that are present in antagonistic strain and absent in target strain (Fig. 5), we further identified the unique BGCs in each antagonistic strain for confronting different targets (Table S9). It has been shown that for bacteria possessing fewer BGCs (e.g., *B. aquimaris* XL39 and *B. horikoshii* ACCC02299), the low number of the unique antibiotics (usually ≤ 3, excluding those without antimicrobial activity, e.g., oligosaccharide and phosphonate), was not appropriate for statistical analysis and can impair the biological regularity of inhibition phenotype against diverse targets (Table S9). This attribution can partially explain the weak correlation between antagonism and phylogeny in these strains. Furthermore, other potential factors may also contribute the relatively low correlation between antagonism and phylogeny, for example, the unknown secondary metabolites of bacteriocins may specifically kill close relatives, internal genetic variation (e.g., point mutations, partial deletion, altered gene regulation, and silent expression) or external cues (e.g.,

bioRxiv preprint doi: https://doi.org/10.1101/2021.08.29.457868; this version posted August 30, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

environmental factors and competing strains) can affect the antibiotic production[11,17,41,42], and the undiscovered genetic and physiological features may also regulate the response to different predicted BGCs. It would be important to identify more secondary metabolites responsible for bacterial interference competition and to further investigate the exquisite regulation characteristics of these functional molecules.

Noticeably, there are some differences between our finding and that reported by Lyons & Kolter, who demonstrated a negative correlation between kin discrimination and phylogeny[14]. This discrepancy may be attributed by the following reasons: (i) The antagonism in this research were evaluated by colony confrontation and fermentation supernatant inhibition, which was dominated by diffusible secondary antibiotics within a comparatively longer distance; while the kin discrimination in Lyons & Kolter's study was assessed based on swarm interaction, biofilm meeting, and halo formation[14], which was likely to be dependent on closer cell association (e.g., toxin-antitoxin system and cell surface contact)[13]. (ii) The antagonism-phylogeny correlation in our study was calculated based on interactions between diverse antagonistic and target strains. However, the halo assay performed previously examined the inhibition of one indicated species (*B. subtilis* NCIB3610) by different testing strains[14]; as discussed above, the different BGCs distribution patterns (e.g., the quantity) among distinct antagonistic strains can influence the inhibition phenotype and its relationship with phylogenetic distance. In general, we consider that Lyons & Kolter's study has provided important knowledge with regard to kin discrimination and close contact, especially in a mixed bacterial population[14]; and our study focused more on the antagonistic interaction *sensu lato*, which can occur within a longer distance.

bioRxiv preprint doi: https://doi.org/10.1101/2021.08.29.457868; this version posted August 30, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Bacilli possess an amazing capacity to synthesize a diverse range of secondary metabolites; previous studies have indicated the phylogenetic conservation of BGCs in the genus *Bacillus* and identified multiple species/clade-specific molecules[16,17,43]. Based on bioinformatics analysis of genomes from a larger scale, we also demonstrated the phylogenetic dependence of BGCs distribution pattern in the genus *Bacillus* (Fig. 2b & 2c). Specifically, the *subtilis* clade appeared to be the most abundant arsenal for secondary metabolites[16], where the 1,259 genomes possessed 16,502 BGCs belonging to 117 products, 310 GCFs, and 47 GCCs, including numerous distinctive and powerful products such as difficidin and macrolactin that produced by *B. amyloliquefaciens* and *B. velezensis*, lichenicidin produced by *B. licheniformis*, and so on (Table S5). On the other hand, the BGC-phylogeny regression suggested that the variation of secondary metabolites among intra- or closely related species was slight and stable, but can be either moderate or drastic among distant groups (Figs. 2c & S5). This finding coincides with the high transferability of BGCs among different genomes and acquisition through horizontal gene transfer (HGT)[16,38], therefore, the correlation (R^2 = 0.2847, Fig. 2c) is relatively lower as compared with other conserved genes (e.g., housekeeping genes or metabolism relevant characteristics)[44].

The cooperation-competition tradeoff among microbial interactions is usually the consequence of the balance between benefits, such as public goods sharing and cross-feeding, and costs, such as resource competition and stress resistance[2,14]. Within a narrow phylogenetic range (e.g., congeneric interactions), only closely associated species can share secreted cooperative goods, such as surfactin for swarming and biofilm formation[14] or siderophores for iron acquisition[45,46], while metabolic similarity decreases moderately or even changes irregularly with phylogenetic distance[9,11].

bioRxiv preprint doi: https://doi.org/10.1101/2021.08.29.457868; this version posted August 30, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Consequently, closely related species can enjoy a higher benefit of public goods than the cost of resource competition and therefore exhibit a relatively weak antagonistic tendency, while distantly related species can hardly share public goods but still confront nutrient competition; thus, the relatively lower benefit than cost contributes to a strong demand for congeneric inhibition. Over a broad phylogenetic range (e.g., intra-phylum interactions), all strains can barely exploit mutually cooperative goods; hence, their benefit-cost balance is predominantly driven by the cost of nutrient competition[9]. As a result, metabolic similarity facilitates the negative correlation between antagonism and phylogenetic distance. This dualistic relationship between interference competition and phylogenetic distance could be a consequence of natural selection, which impels microbes to balance cooperation and competition in an economical manner and plays an important role in regulating community assemblage and succession[47].

In conclusion, the present study demonstrates the consistency between the BGCs distribution and phylogenetic tendency within the genus *Bacillus*, and this coherence acts as the main factor driving the positive correlation between congeneric antagonism and phylogenetic distance, especially in strains possessing abundant BGCs. We expect this positive association between congeneric antagonism and phylogeny is either pronounced[10,11] or can be predicted in other genera with abundant BGCs. This study deepens our understanding of the driving forces and intrinsic mechanism of microbial interactions and provides implications for designing synthetic microbial communities and manipulating population assemblages for practical purposes.

Materials and Methods

bioRxiv preprint doi: https://doi.org/10.1101/2021.08.29.457868; this version posted August 30, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

335 **Bacterial strains and growth conditions**

336 All 90 *Bacillus* strains used in this study are listed in Table S8, including 3 strains obtained from the
337 *Bacillus* Genetic Stock Center (BGSC), 17 strains obtained from the Agricultural Culture Collection of
338 China (ACCC) that originated from different environmental samples, 60 strains isolated from the
339 plant rhizosphere by this laboratory, and 10 mutants of *B. velezensis* SQR9. All strains were grown
340 at 30°C in low-salt Luria-Bertani (LLB) medium (10 g L^{-1} peptone; 5 g L^{-1} yeast extract; 3 g L^{-1} NaCl);
341 when necessary, final concentrations of antibiotics were added as follows: 100 mg L^{-1} spectinomycin
342 (Spc) and 20 mg L^{-1} zeocin (Zeo). To collect the fermentation supernatant for antagonism assessment,
343 the bacterial strains were cultured in Landy medium[48]. The 16S rRNA genes were amplified with the
344 27F (5'-AGAGTTTGATCCTGGCTCAG-3') and 1492R (5'-GGTTACCTTGTTACGACTT-3') primers and
345 were subsequently Sanger sequenced. The taxonomic affiliations of these strains were determined
346 through the EzBioCloud and NCBI databases.

347

348 ***Bacillus* genomic, phylogenetic, and biosynthetic gene cluster analysis**

349 In total, 4,268 available genomes from 139 different *Bacillus* species were downloaded from the
350 NCBI database using the ncbi-genome-download script (https://github.com/kblin/ncbigenome-
351 download/) (Table S1). Then, a phylogenetic tree was constructed based on the concatenation of
352 120 ubiquitous single-copy proteins using GTDB-Tk 1.4.1 software with the default parameters[27,49].
353 The resulting tree was subsequently visualized and edited with Figtree 1.4.4
354 (http://tree.bio.ed.ac.uk/software/figtree/). Maximum likelihood (ML) phylogenetic trees of the 16S
355 rRNA sequence of these strains used in antagonism assessments were constructed by MEGA 5.0.

bioRxiv preprint doi: https://doi.org/10.1101/2021.08.29.457868; this version posted August 30, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Biosynthetic gene clusters (BGCs) of all 4,268 *Bacillus* genomes were predicted using antiSMASH 5.0 software[15]. Considering that numerous congeneric strains shared highly similar BGC profiling, in order to compare the BGC distribution among different *Bacillus* groups more adequately, 545 representative *Bacillus* genomes with high sequencing quality and covering all species, were further selected for BGC distance analysis (Table S1). In detail, the BGC distances were calculated as the weighted combination of the Jaccard Index (JI), adjacency index (AI), and domain sequence similarity (DSS), which resulted in the classification of different gene cluster families (GCFs) and gene cluster clans (GCCs) based on the two cutoff values (0.3 and 0.7, respectively) by the "biosynthetic gene similarity clustering and prospecting engine" (BiG-SCAPE) software[28]. Each *Bacillus* genome was thus annotated to have different GCFs or GCCs, forming a matrix table, which was then organized and visualized using linear models (LM) and hierarchical clustering (HCL) with the average linkage clustering method to view the whole data set by the TIGR multiexperiment viewer (MeV, http://www.tigr.org/software). The BGC distances between different *Bacillus* genomes were then defined as the similarity at the level of GCFs and were directly extracted from the above clustering data. Connection of *Bacillus* genomes with different correlation between BGC distance and phylogenetic distance, was visualized using CIRCOS software (www.circos.ca).

Antagonism assay

The inhibition of antagonistic strains on target strains (listed in Table S8) was evaluated by both colony antagonism and fermentation supernatant inhibition assessments. Five milliliters of a diluted overnight culture of each target strain (~10^5 CFU mL^{-1}) was spread onto LLB plates (10 cm ×10 cm)

bioRxiv preprint doi: https://doi.org/10.1101/2021.08.29.457868; this version posted August 30, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

to be grown as a bacterial lawn. For the colony antagonism assay, 4 μL of the antagonistic strain culture (~10^8 CFU mL^{-1}) was spotted onto plates covered by the target strain layer; for the fermentation supernatant inhibition assay, 150 μL of the filtration-sterilized extracellular supernatant of the antagonistic strain cultured in Landy medium for 48 h was injected into an Oxford cup and then placed on lawns of the target strain. The plates were placed at 22°C until a clear zone formed around the spot, and the inhibition was scored. A heatmap showing the profile of the fermentation supernatant inhibition assay was built by the TIGR multiexperiment viewer (MeV, http://www.tigr.org/software). Each antagonism assay includes three biological replicates.

B. velezensis SQR9 mutant construction

Marker-free deletion strains of the target BGC genes were constructed using the strategy previously described by Zhou et al.[50]. In detail, the left franking (LF) region (~1000 bp), direct repeat (DR) sequence (~500 bp), and right flanking (RF) region (~1000 bp) were amplified from genomic DNA of strain SQR9 using the primer pairs LF-F/LF-R, DR-F/DR-R, and RF-F/RF-R, respectively. The PS cassette (~2300 bp; Spc resistance gene) was amplified with the primers PS-F/PS-R using p7S6 as a template[51]. The four fragments were fused using overlap PCR in the order of LF, DR, PS cassette, and RF. Subsequently, the fused fragments were transformed into competent cells of *B. velezensis* SQR9, and the transformants were selected via first-step screening on LLB plates containing Spc. The final mutants were obtained by combining LLB medium containing 10 mM p-Cl-Phe and the verified primer pair VF/VR. The primers used in this study are listed in Table S10. Each antagonism assay includes three biological replicates.

bioRxiv preprint doi: https://doi.org/10.1101/2021.08.29.457868; this version posted August 30, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

398

399 **Statistics**

400 LM were performed in the R package (version 3.6.1) to assess the correlation between BGC distance
401 and phylogenetic distance, inhibition phenotype and phylogenetic distance/predicted BGC distance,
402 as well as the association of antagonism-phylogeny/BGC-phylogenetic distance/antagonism-BGC
403 distance and the quantity of BGCs in antagonistic strains. Statistical significance was based on a *P*-
404 value < 0.01. Graphs of the correlation data were created using the "ggplot2" package in R; the gray
405 shaded areas denote the 95% confidence intervals. The Duncan's multiple rang tests ($P < 0.05$) of
406 the SPSS version 22.0 (IBM, Chicago, IL, version 22.0) was used for statistical analysis of differences
407 among treatments.

408

409 **Data availability**

410 The authors confirm that the data supporting the findings of this study are available within the article
411 or from the corresponding authors upon request. The DNA sequences from all incubation samples are
412 deposited in the NCBI Sequence GenBank database with accession number listed in Supplementary Table
413 S8.

414

415 **Acknowledgements**

416 This work was financially supported by the National Natural Science Foundation of China (42090060,
417 31972512, and 32072665), the Fundamental Research Funds for the Central Universities
418 (KJQN201919), and the Agricultural Science and Technology Innovation Program of CAAS (CAAS-

bioRxiv preprint doi: https://doi.org/10.1101/2021.08.29.457868; this version posted August 30, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

ZDRW202009).

Author contributions

Q.S., N.Z. and R.Z. conceived and designed the study. L.X. and A.C. performed experimental work. L.X., Y.M., Y.L., Z.L., Z.X. and W.X. analyzed the data. Y.L., X.S. and Y.X. contributed materials. N.Z., R.Z., Y.M. and L.X. wrote the manuscript.

Conflict of Interest

The authors declare no conflict of interest.

References

1. Pérez-Gutiérrez, R.-A. *et al.* Antagonism influences assembly of a *Bacillus* guild in a local community and is depicted as a food-chain network. *ISME J* **7**, 487-497 (2013).

2. Ghoul, M. & Mitri, S. The ecology and evolution of microbial competition. *Trends Microbiol* **24**, 833-845 (2016).

3. Hibbing, M.E., Fuqua, C., Parsek, M.R. & Peterson, S.B. Bacterial competition: surviving and thriving in the microbial jungle. *Nat Rev Microbiol* **8**, 15-25 (2010).

4. Granato, E.T., Meiller-Legrand, T.A. & Foster, K.R. The evolution and ecology of bacterial warfare. *Curr Biol* **29**, R521-R537 (2019).

5. Gonzalez, D., Sabnis, A., Foster, K.R. & Mavridou, D.A.I. Costs and benefits of provocation in bacterial warfare. *Proc Natl Acad Sci U S A* **115**, 7593-7598 (2018).

6. Jamet, A. & Nassif, X. New players in the toxin field: polymorphic toxin systems in bacteria. *mBio* **6**, e00285-15 (2015).

7. Cordero, O.X. *et al.* Ecological populations of bacteria act as socially cohesive units of antibiotic

bioRxiv preprint doi: https://doi.org/10.1101/2021.08.29.457868; this version posted August 30, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

production and resistance. *Science* **337**, 1228-31 (2012).

8. P., d.V.R. *et al.* Comparative genomics reveals high biological diversity and specific adaptations in the industrially and medically important fungal genus *Aspergillus*. *Genome Biol* **18**, 28 (2017).

9. Russel, J., Røder, H.L., Madsen, J.S., Burmølle, M. & Sørensen, S.J. Antagonism correlates with metabolic similarity in diverse bacteria. *Proc Natl Acad Sci U S A* **114**, 10684-10688 (2017).

10. Vetsigian, K., Jajoo, R. & Kishony, R. Structure and evolution of *Streptomyces* interaction networks in soil and in silico. *PLoS Biol* **9**, e1001184 (2011).

11. Westhoff, S., Kloosterman, A., van Hoesel, S.F.A., van Wezel, G.P. & Rozen, D.E. Competition sensing alters antibiotic production in *Streptomyces*. *mBio* **12**, e02729-20 (2021).

12. Stefanic, P., Kraigher, B., Lyons, N.A., Kolter, R. & Mandic-Mulec, I. Kin discrimination between sympatric *Bacillus subtilis* isolates. *PNAS* **112**, 14042-14047 (2015).

13. Lyons, N.A., Kraigher, B., Stefanic, P., Mandic-Mulec, I. & Kolter, R. A combinatorial kin discrimination system in *Bacillus subtilis*. *Curr Biol* **26**, 733-742 (2016).

14. Lyons, N.A. & Kolter, R. *Bacillus subtilis* protects public goods by extending kin discrimination to closely related species. *mBio* **8**, e00723-17 (2017).

15. Blin, K. *et al.* antiSMASH 5.0: updates to the secondary metabolite genome mining pipeline. *Nucleic Acids Res* **47**, W81-W87 (2019).

16. Steinke, K., Mohite, O.S., Weber, T. & Kovacs, A.T. Phylogenetic distribution of secondary metabolites in the *Bacillus subtilis* species complex. *mSystems* **6**, e00057-21 (2021).

17. Kiesewalter, H.T. *et al.* Genomic and chemical diversity of *Bacillus subtilis* secondary metabolites against plant pathogenic fungi. *mSystems* **6**, e00770-20 (2021).

bioRxiv preprint doi: https://doi.org/10.1101/2021.08.29.457868; this version posted August 30, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

18. Wang, D. *et al.* A genomic island in a plant beneficial rhizobacterium encodes novel antimicrobial fatty acids and a self-protection shield to enhance its competition. *Environ Microbiol* (2019).

19. Chowdhury, S.P., Hartmann, A., Gao, X. & Borriss, R. Biocontrol mechanism by root-associated *Bacillus amyloliquefaciens* FZB42 - a review. *Front Microbiol* **6**, 780 (2015).

20. Gross, H. & Loper, J.E. Genomics of secondary metabolite production by *Pseudomonas* spp. *Nat Prod Rep* **26**, 1408-1446 (2009).

21. Chen, X. *et al.* Structural and functional characterization of three polyketide synthase gene clusters in *Bacillus amyloliquefaciens* FZB 42. *J Bacteriol* **188**, 4024-4036 (2006).

22. Chen, X.H. *et al.* Comparative analysis of the complete genome sequence of the plant growth-promoting bacterium *Bacillus amyloliquefaciens* FZB42. *Nat Biotechnol* **25**, 1007-14 (2007).

23. Ongena, M. & Jacques, P. *Bacillus* lipopeptides: versatile weapons for plant disease biocontrol. *Trends Microbiol* **16**, 115-125 (2008).

24. Sansinenea, E. & Ortiz, A. Secondary metabolites of soil *Bacillus* spp. *Biotechnol Lett* **33**, 1523-38 (2011).

25. Zhao, X. & Kuipers, O.P. Identification and classification of known and putative antimicrobial compounds produced by a wide variety of Bacillales species. *BMC Genomics* **17**, 882 (2016).

26. Letzel, A., Pidot, S.J. & Hertweck, C. Genome mining for ribosomally synthesized and post-translationally modified peptides (RiPPs) in anaerobic bacteria. *BMC Genomics* **15**, 983 (2014).

27. Parks, D.H. *et al.* A standardized bacterial taxonomy based on genome phylogeny substantially revises the tree of life. *Nat Biotechnol* **36**, 996-1004 (2018).

28. Navarro-Munoz, J.C. *et al.* A computational framework to explore large-scale biosynthetic diversity.

bioRxiv preprint doi: https://doi.org/10.1101/2021.08.29.457868; this version posted August 30, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Nat Chem Biol **16**, 60-68 (2020).

29. Wilson, D.N. Ribosome-targeting antibiotics and mechanisms of bacterial resistance. *Nat Rev Microbiol* **12**, 35-48 (2014).

30. Zhang, N. *et al.* Whole transcriptomic analysis of the plant-beneficial rhizobacterium *Bacillus amyloliquefaciens* SQR9 during enhanced biofilm formation regulated by maize root exudates. *BMC Genomics* **16**, 685 (2015).

31. Xu, Z. *et al.* Enhanced control of cucumber wilt disease by *Bacillus amyloliquefaciens* SQR9 by altering the regulation of its DegU phosphorylation. *Appl Environ Microbiol* **80**, 2941-2950 (2014).

32. Darwin, C. On the Origin of Species. *John Murray, London* (1859).

33. Alexandrou, M.A. *et al.* Evolutionary relatedness does not predict competition and co-occurrence in natural or experimental communities of green algae. *Proc Biol Sci* **282**, 20141745 (2015).

34. Kunstler, G. *et al.* Plant functional traits have globally consistent effects on competition. *Nature* **529**, 204-7 (2016).

35. Venail, P.A. *et al.* The influence of phylogenetic relatedness on species interactions among freshwater green algae in a mesocosm experiment. *J Ecol* **102**, 1288-1299 (2014).

36. Violle, C., Nemergut, D.R., Pu, Z. & Jiang, L. Phylogenetic limiting similarity and competitive exclusion. *Ecol Lett* **14**, 782-7 (2011).

37. van Bergeijk, D.A., Terlouw, B.R., Medema, M.H. & van Wezel, G.P. Ecology and genomics of Actinobacteria: new concepts for natural product discovery. *Nat Rev Microbiol* **18**, 546-558 (2020).

38. Khaldi, N., Collemare, C., Lebrun, M. & Wolfe, K.H. Evidence for horizontal transfer of a secondary metabolite gene cluster between fungi. *Genome Biol* **9**, R18 (2008).

bioRxiv preprint doi: https://doi.org/10.1101/2021.08.29.457868; this version posted August 30, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

503 39. Crits-Christoph, A., Diamond, S., Butterfield, C.N., Thomas, B.C. & Banfield, J.F. Novel soil bacteria
504 possess diverse genes for secondary metabolite biosynthesis. *Nature* **558**, 440-444 (2018).
505 40. Dittmann, E., Gugger, M., Sivonen, K. & Fewer, D.P. Natural product biosynthetic diversity and
506 comparative genomics of the Cyanobacteria. *Trends Microbiol* **23**, 642-652 (2015).
507 41. Yun, C.S., Motoyama, T. & Osada, H. Biosynthesis of the mycotoxin tenuazonic acid by a fungal NRPS-
508 PKS hybrid enzyme. *Nat Commun* **6**, 8758 (2015).
509 42. Zhang, M.M. *et al.* CRISPR-Cas9 strategy for activation of silent *Streptomyces* biosynthetic gene
510 clusters. *Nat Chem Biol* **13**, 607-609 (2017).
511 43. Grubbs, K.J. *et al.* Large-scale bioinformatics analysis of *Bacillus* genomes uncovers conserved roles
512 of natural products in bacterial physiology. *mSystems* **2**, e00040-17 (2017).
513 44. Levy, R. & Borenstein, E. Metabolic modeling of species interaction in the human microbiome
514 elucidates community-level assembly rules. *Proc Natl Acad Sci U S A* **110**, 12804-12809 (2013).
515 45. Griffin, A.S., West, S.A. & Buckling, A. Cooperation and competition in pathogenic bacteria. *Nature*
516 **430**, 1024-1027 (2004).
517 46. Gu, S. *et al.* Competition for iron drives phytopathogen control by natural rhizosphere microbiomes.
518 *Nat Microbiol* **5**, 1002-1010 (2020).
519 47. Zhou, J. & Ning, D. Stochastic community assembly: does it matter in microbial ecology? *Microbiol*
520 *Mol Biol R* **81**, e00002-17 (2017).
521 48. Landy, M., Rosenman, S.B. & Warren, G.H. An antibiotic from *Bacillus subtilis* active against
522 pathogenic fungi. *J Bacteriol* **54**, 24 (1947).
523 49. Chaumeil, P.A., Mussig, A.J., Hugenholtz, P. & Parks, D.H. GTDB-Tk: a toolkit to classify genomes with

bioRxiv preprint doi: https://doi.org/10.1101/2021.08.29.457868; this version posted August 30, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

the Genome Taxonomy Database. *Bioinformatics* (2019).

50. Zhou, C. *et al.* pheS* , an effective host-genotype-independent counter-selectable marker for marker-free chromosome deletion in *Bacillus amyloliquefaciens*. *Appl Microbiol Biotechnol* **101**, 217-227 (2017).

51. Yan, X., Yu, H.J., Hong, Q. & Li, S.P. Cre/lox system and PCR-based genome engineering in *Bacillus subtilis*. *Appl Environ Microbiol* **74**, 5556-62 (2008).

bioRxiv preprint doi: https://doi.org/10.1101/2021.08.29.457868; this version posted August 30, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Figures

Fig. 1

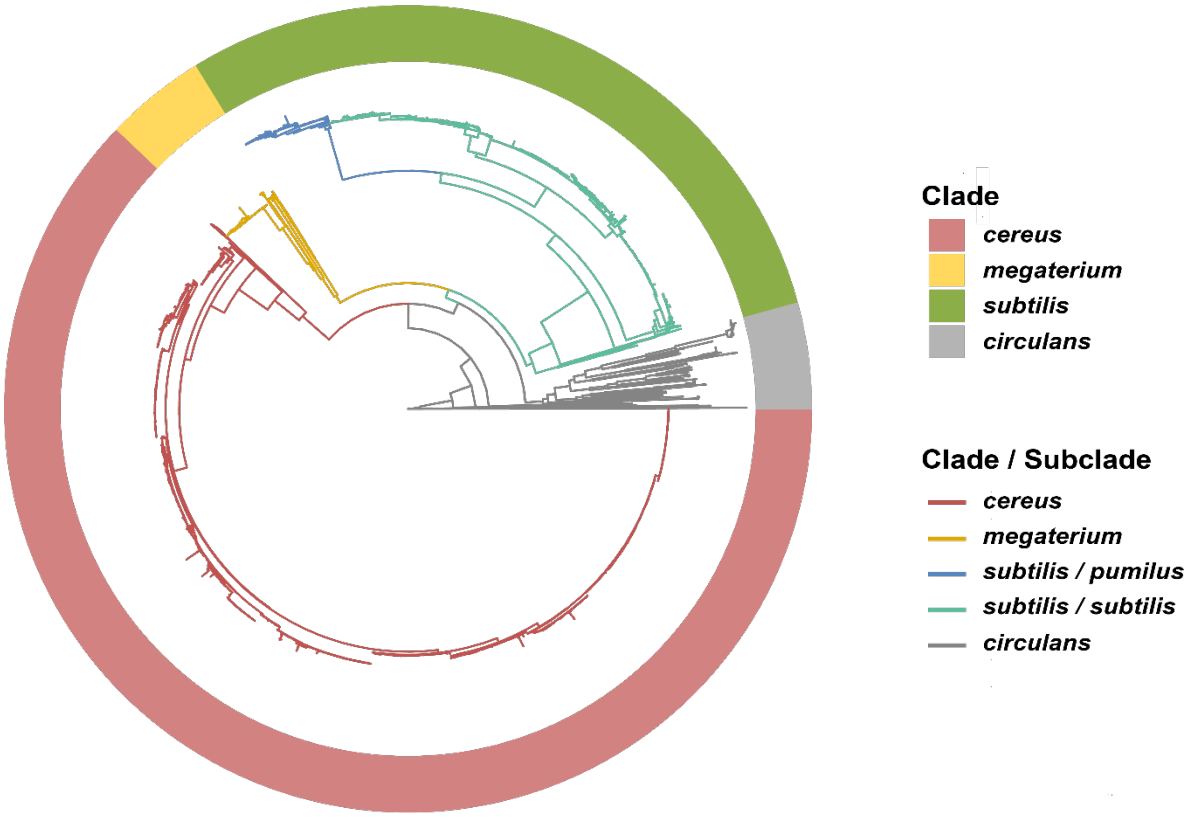
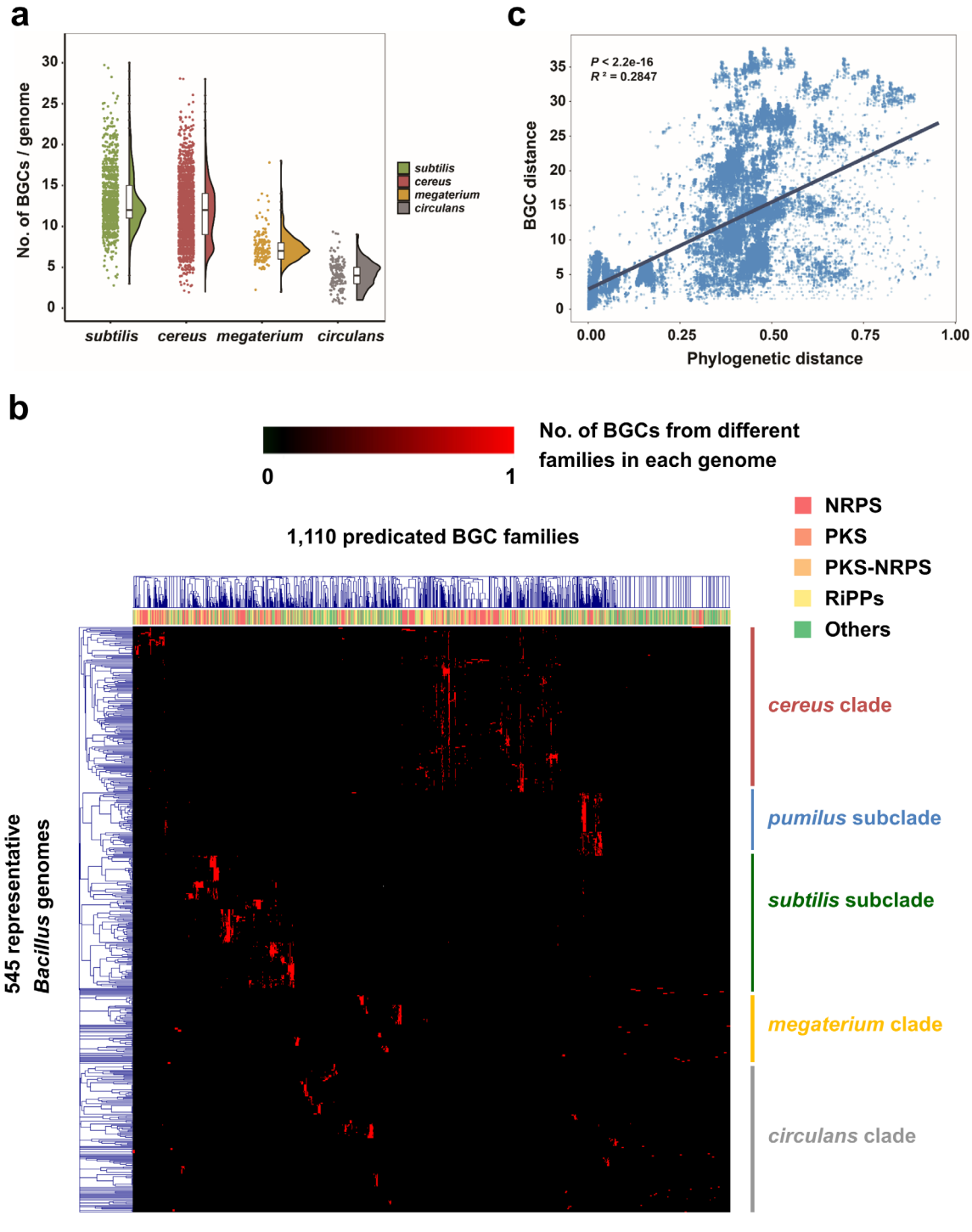


Fig. 1 Phylogram of the tested *Bacillus* genomes. The maximum likelihood (ML) phylogram of 4,268 *Bacillus* genomes was based on the sequences of 120 ubiquitous single-copy proteins[27]. The phylogenetic tree shows that *Bacillus* species can be generally clustered into the *subtilis* (light green circle; further includes *subtilis* (dark green) and *pumilus* (blue) subclades as shown in the branches), *cereus* (red), *megaterium* (yellow), and *circulans* (gray) clades. For detailed information of the species, please refer to the phylogenetic tree in Fig. S1.

bioRxiv preprint doi: https://doi.org/10.1101/2021.08.29.457868; this version posted August 30, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

539 **Fig. 2**



540

541 **Fig. 2 Biosynthetic gene cluster (BGC) distribution is correlated with phylogeny in the genus *Bacillus*. (a)** The

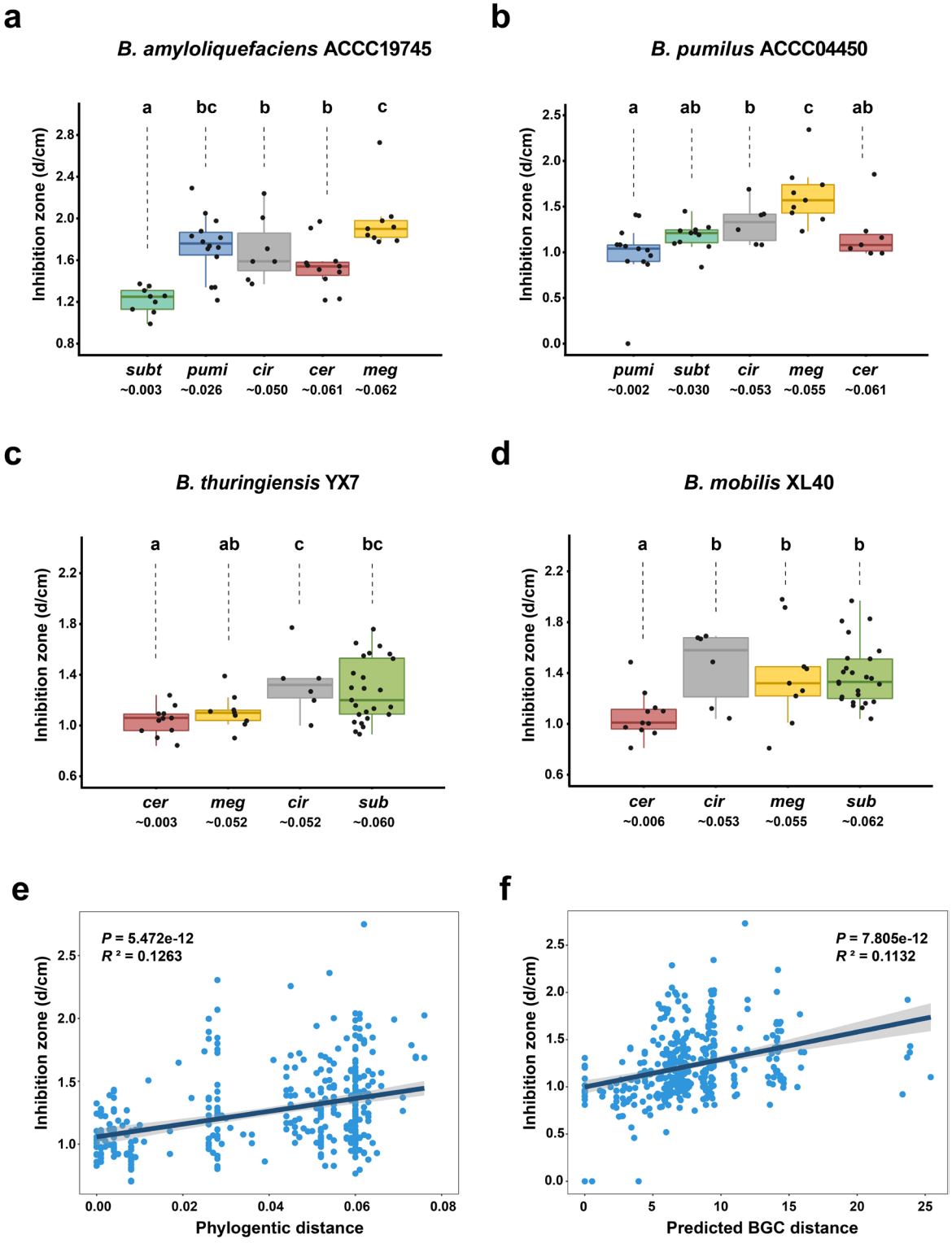
542 numbers of BGCs in the 4,268 *Bacillus* genomes from different clades as defined by antiSMASH[15]. **(b)**

bioRxiv preprint doi: https://doi.org/10.1101/2021.08.29.457868; this version posted August 30, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Hierarchal clustering among the 545 representative *Bacillus* genomes based on the abundance of the different biosynthesis gene cluster families (GCFs). Each row represents a GCF, which was classified through BiG-SCAPE by calculating the Jaccard index (JI), adjacency index (AI), and domain sequence similarity (DSS) of each BGC[28]; the color bar on the top of the heatmap represents the BGC class of each GCF, where PKS includes classes of PKSother and PKSI, PKS-NRPS means PKS-NRPS Hybrids, Others includes classes of saccharides, terpene, and others. Each line represents a *Bacillus* genome, and the abundance of each GCF in different genomes is shown in the heatmap. The left tree was constructed based on the distribution pattern of GCFs, which showed a similar pattern to the phylogram in **Fig. 1**. **(c)** The correlation between the BGC and phylogenetic distance of the 545 representative *Bacillus* genomes ($P < 0.001$, $R^2 = 0.2847$).

bioRxiv preprint doi: https://doi.org/10.1101/2021.08.29.457868; this version posted August 30, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

552 **Fig. 3**



553

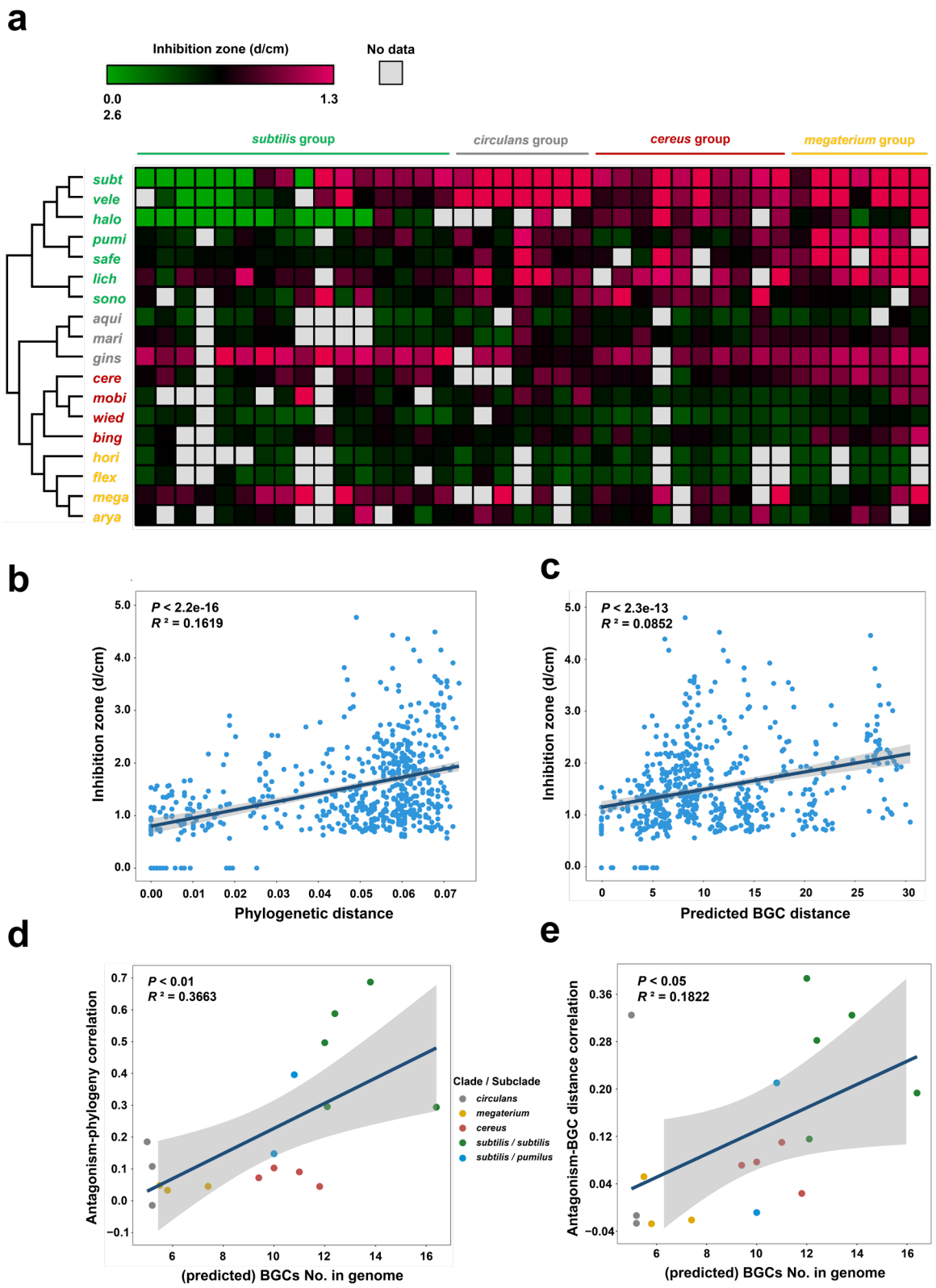
554 **Fig. 3 Colony antagonism phenotype is positively correlated with the phylogenetic and BGC distance within**

bioRxiv preprint doi: https://doi.org/10.1101/2021.08.29.457868; this version posted August 30, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

***Bacillus* species. (a~d)** Inhibition of colonies of *B. amyloliquefaciens* ACCC19745, *B. pumilus* ACCC04450, *B. thuringiensis* YX7, and *B. mobilis* XL40 against *Bacillus* from different clades. The number below the abbreviations indicates the average 16S rRNA gene phylogenetic distance of the target strains with the corresponding antagonistic strain. Each inhibition assay includes three biological replicates and the average is shown in the boxplots and used for the correlation analysis; boxplots with different letters are statistically different according to the Duncan's multiple range tests ($P < 0.05$). **(e & f)** Correlation between the antagonism phenotype (diameter of the inhibition zone) and 16S rRNA phylogenetic **(e)** or predicted BGC distance **(f)** among all the tested paired strains. For strains whose genomes have not been completely sequenced, we referred to the *Bacillus* genomes in the NCBI database that shared the highest 16S rRNA similarity.

bioRxiv preprint doi: https://doi.org/10.1101/2021.08.29.457868; this version posted August 30, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

564 **Fig. 4**



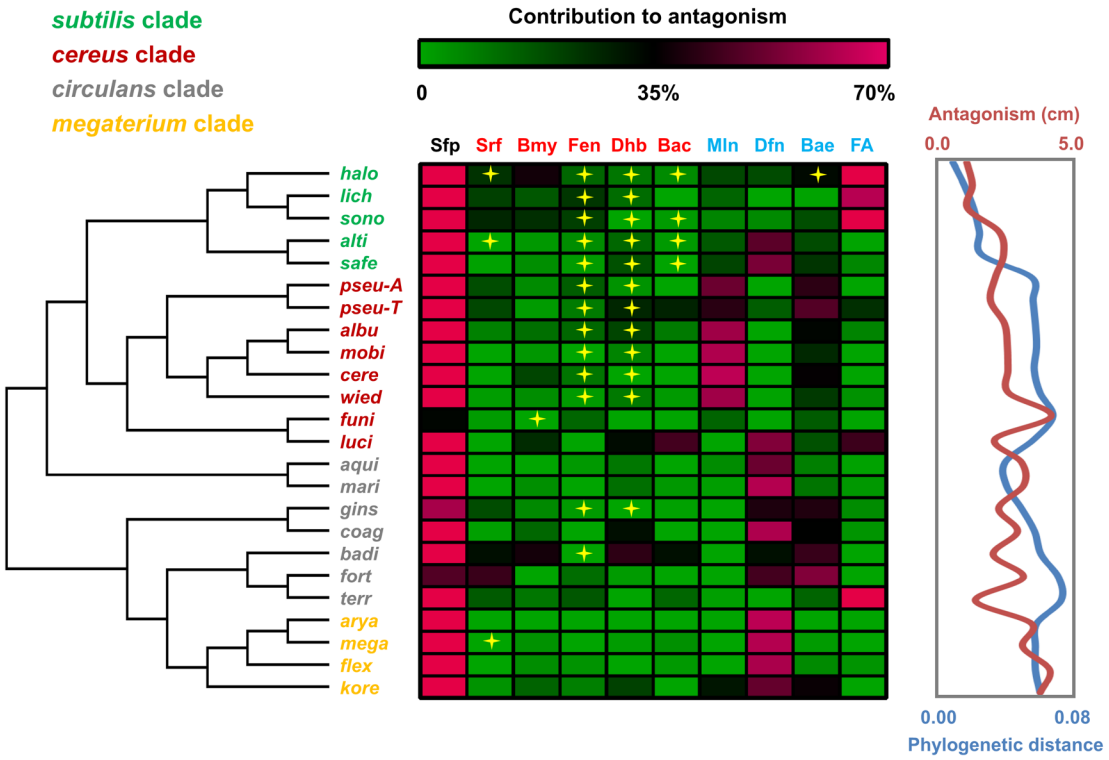
565

bioRxiv preprint doi: https://doi.org/10.1101/2021.08.29.457868; this version posted August 30, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

566 **Fig. 4 Congeneric inhibition by fermentation supernatants is positively correlated with the phylogenetic and**
567 **BGC distance in *Bacillus*. (a)** Heatmap showing the antagonistic profiles of the fermentation supernatant of
568 17 antagonistic strains (in the left column) on the 40 target strains (in the top line). The maximum likelihood
569 (ML) phylogenetic tree was constructed based on the 16S rRNA sequence of the 17 antagonistic strains: *subt*,
570 *B. subtilis* RZ30; *vele*, *B. velezensis* SQR9; *halo*, *B. halotolerans* CF7; *pumi*, *B. pumilus* ACCC04450; *safe*, *B.*
571 *safensis* LY9; *lich*, *B. licheniformis* CC11; *sono*, *B. sonorensis* YX13; *aqui*, *B. aquimaris* XL39; *mari*, *B. marisflavi*
572 XL37; *gins*, *B. ginsengihumi* ACCC05679; *cere*, *B. cereus* ACCC10263; *mobi*, *B. mobilis* XL40; *wied*, *B. wiedmannii*
573 XL36; *bing*, *B. bingmayongensis* KF27; *hori*, *B. horikoshii* ACCC02299; *flex*, *B. flexus* DY11; *mega*, *B. megaterium*
574 ACCC01509. Each inhibition assay includes three biological replicates and the average is shown in the heatmap
575 and used for the correlation analysis. **(b & c)** Correlation between the antagonism phenotype (diameter of the
576 inhibition zone) and 16S rDNA phylogenetic distance **(b)** or the predicted BGC distance **(c)** among all the tested
577 paired strains. For strains whose genomes have not been completely sequenced, we referred to the *Bacillus*
578 genomes in the NCBI database that shared the highest 16S rRNA similarity. **(d & e)** Correlation between the
579 antagonism-phylogeny association **(d)** or antagonism-BGC distance association **(e)** and the (predicted)
580 quantity of BGCs in antagonistic strains. For strains whose genomes have not been completely sequenced, we
581 referred to the average quantity of BGCs in this species. The color of the dots represents the clade/subclade
582 which the antagonistic strains belong to.

bioRxiv preprint doi: https://doi.org/10.1101/2021.08.29.457868; this version posted August 30, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

583 **Fig. 5**



584

585 **Fig. 5 Contribution of BGCs to antagonizing *Bacillus* species from different clades by *B. velezensis* SQR9.** The

586 heatmap shows the contribution of each BGC product (on the top) to the inhibition of each target strain (on

587 the left), which was calculated as the percentage of the decreased inhibition zone of the corresponding BGC-

588 deficient mutant compared with wild type. The maximum likelihood (ML) tree on the left was constructed

589 based on the 16S rRNA sequences of the 24 target strains: *halo*, *B. halotolerans* CF7; *lich*, *B. licheniformis* LY2;

590 *sono*, *B. sonorensis* YX13; *alti*, *B. altitudinis* LY37; *safe*, *B. safensis* LY9; *pseu*-A, *B. pseudomycoides* ACCC10238;

591 *pseu*-B, *B. pseudomycoides* TZ8; *albu*, *B. albus* XL388; *mobi*, *B. mobilis* XL40; *cere*, *B. cereus* ACCC10263; *wied*,

592 *B. wiedmannii* CF23; *funi*, *B. funiculus* ACCC05674; *luci*, *B. luciferensis* XL165; *aqui*, *B. aquimaris* XL39; *mari*, *B.*

593 *marisflavi* XL37; *gins*, *B. ginsengihumi* ACCC05679; *coag*, *B. coagulans* ACCC10229; *badi*, *B. badius* ACCC60106;

594 *fort*, *B. fortis* ACCC10219; *terr*, *B. terrae* TL19; *arya*, *B. aryabhattai* XL26; *mega*, *B. megaterium* ACCC01509;

bioRxiv preprint doi: https://doi.org/10.1101/2021.08.29.457868; this version posted August 30, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

595 *flex*, *B. flexus* DY11; *kore, B. koreensis* ACCC05681. The abbreviations on the top (except Sfp) represent
596 different BGC products: Srf, surfactin; Bmy, bacillomycin D; Fen, fengycin; Dhb, bacillibactin; Bac, bacilysin;
597 Mln, macrolactin; Bae, bacillaene; Dfn, difficidin; FA, an antimicrobial fatty acid, bacillunoic acid. Specifically,
598 Sfp (phospho-pantheinyltransferase) is not an antibiotic but is necessary for modification of the above
599 antibiotics and ensuring their activity, except for bacilysin[22,30]; here, the contribution of Sfp to antagonism
600 means the relative reduction of inhibition by SQR9Δ*sfp* compared to that of wild-type SQR9 towards different
601 targets. The curves in the right box show the antagonistic phenotype and phylogenetic distance of *B. velezensis*
602 SQR9 with each of the target strains. Each inhibition assay included three biological replicates and the average
603 contribution was shown in the heatmap.

bioRxiv preprint doi: https://doi.org/10.1101/2021.08.29.457868; this version posted August 30, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Supporting Information

Table S1 Information of the 4,268 *Bacillus* genomes used for the phylogenetic analysis and biosynthetic gene cluster (BGC) prediction.

Table S2 Taxonomic distribution of genomes across different species and clades.

Table S3 Prediction of the putative biosynthetic gene clusters (BGCs) in all 4,268 *Bacillus* genomes by using antiSMASH.

Table S4 Statistics of different classes of biosynthetic gene clusters (BGCs) in different *Bacillus* clades.

Table S5 Biosynthesis gene cluster families (GCFs) and gene cluster clans (GCCs) in the 545 representative *Bacillus* genomes based on interactive sequence similarity network analysis using BiG-SCAPE.

Table S6 Product information of each predicted biosynthesis gene cluster family (GCF) and gene cluster clan (GCC) in the 545 representative *Bacillus* genomes.

Table S7 Statistics of biosynthetic gene clusters families (GCFs) in the 545 representative *Bacillus* genomes.

Table S8 Bacterial strains used in this study.

Table S9 Quantity of specific biosynthetic gene clusters (BGCs) of each antagonistic strain when confronted with different target strains in the fermentation supernatant inhibition assay. For strains whose genome have not been completely sequenced, the BGC presence was assigned if more than 80% of the corresponding *Bacillus* species genomes possessed this cluster, and their total predicted BGCs No. was calculated as the average BGCs. No. in all genomes of this species.

Table S10 Primers used for mutants construction and verification.

bioRxiv preprint doi: https://doi.org/10.1101/2021.08.29.457868; this version posted August 30, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Fig. S1 Maximum likelihood (ML) phylogenetic tree of the 4,268 *Bacillus* genomes based on the sequences of 120 ubiquitous single-copy proteins[27]. This phylogram is the fully-annotated version of Fig. 1, which shows the detailed species information of the *Bacillus* genomes.

Fig. S2 Distribution of the biosynthetic gene clusters (BGCs) types. BGCs belonging to PKSI or EPS are not shown in this figure since they are extremely rare in *Bacillus* genomes.

Fig. S3 Profile of biosynthetic gene cluster (BGC) products and classification attached to the phylogenetic tree of total 4,268 *Bacillus* genomes. Matching to the genome order in the phylogenetic tree (the same as that in Figs. 1 & S1), BGCs profile in each genome were shown as the number of 256 detailed product types (Table S3) by the heatmap (which were further clustered only at the level of BGC products), as well as the number of 8 different classifications (NPRS, PKSI, PKSother, PKS-NRP_Hybrids, RiPPs, Saccharides, Terpene, and Others) through the histogram.

Fig. S4 Hierarchal clustering among the 545 representative *Bacillus* genomes based on the abundance of the different biosynthesis gene cluster families (GCFs). This figure is the fully-annotated version of Fig. 2b. Each row represents a GCF and has been noted with the potential BGC product, which was classified through BiG-SCAPE by calculating the Jaccard index (JI), adjacency index (AI), and domain sequence similarity (DSS) of each BGC[28]. Each line represents a *Bacillus* genome, and the abundance of each GCF in different genomes is shown in the heatmap. The left tree was constructed based on the distribution pattern of BGCs from different families, which showed a similar pattern to the phylogram in **Fig. 1**.

Fig. S5 Connection of *Bacillus* genomes in 7 groups showing different correlation between biosynthetic gene clusters (BGCs) distance and phylogenetic distance. (a) The correlation between

bioRxiv preprint doi: https://doi.org/10.1101/2021.08.29.457868; this version posted August 30, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

the BGC and phylogenetic distance of the 545 representative *Bacillus* genomes forms 7 distinguishable groups (group 1~7). **(b~h)** For detail analysis, the points of 7 groups in **(a)**, corresponding to different relationships of two *Bacillus* genomes, were extracted, respectively, and shown as the lines, which connected each two different genomes indicated by their own points. All connected genomes were ranked as the orders in phylogenetic tree of the 545 *Bacillus* genomes from top to bottom, as that shown in **(b)**. Different colors were used to indicate each *Bacillus* clades/subclades: red, *cereus* clade; blue, *pumilus* subclade; green, *subtilis* subclade; yellow, *megaterium* clade; gray, *circulans* clade. Pd: phylogenetic distance; BGCd: BGC distance.

Fig. S6 Visual representations of colony confrontation and fermentation supernatant inhibition assay.

Fig. S7 Correlation of the BGC-phylogenetic distance association with the (predicted) quantity of BGCs in antagonistic strains. For strains whose genomes have not been completely sequenced, we referred to the average quantity of BGCs in this species. The color of the dots represents the clade/subclade which the antagonistic strains belong to.