Phylogeny Reveals Novel HipA-Homologous Kinase Families and Toxin-Antitoxin Gene Organizations

Gerdes, Kenn; Bærentsen, Rene; Brodersen, Ditlev E.

Published in:
mBio

DOI:
10.1128/mBio.01058-21

Publication date:
2021

Document version
Publisher's PDF, also known as Version of record

Document license:
CC BY

Citation for published version (APA):
https://doi.org/10.1128/mBio.01058-21
Phylogeny Reveals Novel HipA-Homologous Kinase Families and Toxin-Antitoxin Gene Organizations

Kenn Gerdes,a Rene Bærentzen,b Ditlev E. Brodersenb

aCentre of Excellence for Bacterial Stress Response and Persistence, Section for Functional Genomics, Department of Biology, University of Copenhagen, Copenhagen N, Denmark
bDepartment of Molecular Biology and Genetics, Aarhus University, Aarhus C, Denmark

ABSTRACT
Toxin-antitoxin modules function in the genetic stability of mobile genetic elements, bacteriophage defense, and antibiotic tolerance. A gain-of-function mutation of the Escherichia coli K-12 hipBA module can induce antibiotic tolerance in a subpopulation of bacterial cells, a phenomenon known as persistence. HipA is a Ser/Thr kinase that phosphorylates and inactivates glutamyl tRNA synthetase, inhibiting cellular translation and inducing the stringent response. Additional characterized HipA homologues include HipT from pathogenic E. coli O127 and YjjJ of E. coli K-12, which are encoded by tricistronic hipBST and monocistronic operons, respectively. The apparent diversity of HipA homologues in bacterial genomes inspired us to investigate overall phylogeny. Here, we present a comprehensive phylogenetic analysis of the Hip kinases in bacteria and archaea that expands on this diversity by revealing seven novel kinase families. Kinases of one family, encoded by monocistronic operons, consist of an N-terminal core kinase domain, a HipS-like domain, and a HIRAN (HIP116 Rad5p N-terminal) domain. HIRAN domains bind single- or double-stranded DNA ends. Moreover, five types of bicistronic kinase operons encode putative antitoxins with HipS-HIRAN, HipS, γ5-resolvase, or Stl repressor-like domains. Finally, our analysis indicates that reversion of hipBA gene order happened independently several times during evolution.

IMPORTANCE Bacterial multidrug tolerance and persistence are problems of increasing scientific and medical significance. The first gene discovered to confer persistence was hipA, encoding the kinase toxin of the hipBA toxin-antitoxin (TA) module of E. coli. HipA-homologous kinases phosphorylate and thereby inactivate specific tRNA synthetases, thus inhibiting protein translation and cell proliferation. Here, we present a comprehensive phylogenetic analysis of bacterial Hip kinases and discover seven new families with novel operon structures and domains. Overall, Hip kinases are encoded by TA modules with at least 10 different genetic organizations, 7 of which have not been described before. These results open up exciting avenues for the experimental analysis of the superfamily of Hip kinases.

KEYWORDS high persister A, HipB, HipS, HipT, HIRAN, Stl, GltX, TrpS, kinase

Prokaryotic toxin-antitoxin (TA) modules were discovered due to their ability to stabilize plasmids by killing of plasmid-free cells by a mechanism known as postsegregational killing (1, 2). The mechanism relies on stable protein toxins that are inhibited either by unstable antitoxin RNAs (type I and III TAs) or unstable antitoxin proteins (type II TAs) as long as the plasmid remains in the cell. If, on the other hand, the plasmid is lost, degradation of antitoxin leads to toxin activation and hence, death of the plasmid-free cell. Since their discovery on plasmids, TAs have been identified on a wide range of bacterial and archael chromosomes as well (3–5), often in multiple or even large numbers (5–9). For example, Mycobacterium tuberculosis carries genes that

Editor Michael T. Laub, Massachusetts Institute of Technology
Copyright © 2021 Gerdes et al. This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International license.
Address correspondence to Kenn Gerdes, kenn.gerdes5@gmail.com, or Ditlev E. Brodersen, deb@mbg.au.dk.
This article is a direct contribution from Kenn Gerdes, a Fellow of the American Academy of Microbiology, who arranged for and secured reviews by Pierre Genevaux, CNRS-University Paul-Sabatier, and Laurent Falquet, University of Fribourg and Swiss Institute of Bioinformatics.
Received 14 April 2021
Accepted 20 April 2021
Published 1 June 2021
encode ~70 type II TA modules (7, 10, 11), while the plant symbiont Sinorhizobium meliloti contains more than 100 (12). The biological functions of chromosome-encoded TAs are debated, but experimental evidence supports at least three roles that are not mutually exclusive: (i) genetic stabilization of chromosome segments or entire chromosomes (13–16), (ii) antiphage defense by abortive infection (17, 18), and (iii) antibiotic tolerance (19–22). Intriguingly, it was recently discovered that bacterial retrons encode a type of three-component TAs that can function as antiphage defense systems, thus supporting the notion that a major function of TAs may in fact be to curb or control bacteriophage infection (23–26). This idea is consistent with the recent observation that environmental or nutritional stress in general does not activate type II TA-encoded toxins (27).

There is evidence supporting that some TAs induce antibiotic tolerance or persistence in bacteria. Persistence is a phenomenon found in all bacteria tested (19, 28–30) and is operationally defined as the subpopulation of a bacterial cells that survive for an extended period of time in the presence of inhibitory concentrations of antibiotics (31). Common to persistence mechanisms is that the phenotype is a stochastic phenomenon and expressed by only a fraction of the cell population at any given time (31, 32). Importantly, persistence is believed to contribute to the recalcitrance of bacterial infections and may thus pose a significant medical problem (30, 33–35). At the mechanistic level, persisters are slow-growing cells that display increased survival rates in the presence of antibiotics (32, 36, 37). In addition, this also buys the bacterial population time to develop true antibiotic resistance (38). Apart from a reduced growth rate, persister cells can also arise from expression of high levels of factors that counter the effects of antibiotics in a small subpopulation of cells (39, 40).

HipA (high persister gene A) of Escherichia coli K-12 was the first gene found to be associated with increased persistence based on the identification of the gain-of-function allele hipA7 in a strain exhibiting increased tolerance toward penicillin (41). The mutant allele, later also found in clinical isolates of uropathogenic E. coli (28), exhibits a 100- to 1,000-fold increased level of persistence (32, 42), but even the wild-type hipBA module can be shown to confer a modest, but measurable, increase of persistence (28). The hipA toxin gene and its upstream hipB antitoxin gene constitute a canonical type II TA module encoding two proteins that combine to form an inactive HipBA complex, which, upon degradation of HipB, generates active HipA toxin (Fig. 1A) (43). Consequently, ectopic production of HipA in E. coli causes severe growth inhibition that can be reversed by later expression of HipB antitoxin (43). HipBA from both E. coli K-12 and Shewanella oneidensis MR-1 assemble into heterotetrameric HipA2B2 complexes (28, 44–48).

HipA is a so-called Hanks serine-threonine kinase (49, 50) that phosphorylates a conserved serine residue (Ser239) in glutamyl-tRNA synthetase (GltX) inside its bacterial host, inhibiting the enzyme and thereby aminoacylation of tRNA_Glu (51, 52). As a consequence, the ratio of charged to uncharged tRNA_Glu decreases, thus stimulating RelA-tRNA_Glu complexes to bind the ribosomal A site. Activation of RelA (53) on the ribosome leads to an increase in the cellular (p)ppGpp level triggering the stringent response (42, 51, 54). HipA shares its fold with human cyclin-dependent kinases and maintains all of the conserved motifs necessary for kinase activity (45). The antitoxin HipB contains a classical (Cro-like) helix-turn-helix (HTH) DNA-binding domain (45) and forms a homodimer when in complex with HipA that allows binding to palindromic operator sequences in the hipBA promoter region (Fig. 1A). The mechanism of toxin inhibition by HipB has not been fully elucidated, but it appears to differ from most type II systems in that the antitoxin does not interact directly with the toxin active site. Instead, the very C terminus of HipB appears to bind in a pocket on HipA, possibly regulating toxin activity (44). Finally, HipA can inactivate its own kinase activity by auto-phosphorylation (55), a phenomenon that has been proposed to function in resuscitation of persister cells (46, 56).

HipA homologues are present in many bacteria. For example, the chromosome of the Alphaproteobacterium Caulobacter crescentus contains three hipBA loci for which
the encoded HipA toxins inhibit protein synthesis upon ectopic expression in all three cases (57). Like their *E. coli* counterpart, HipA1 and HipA2 contribute to persistence during stationary phase by phosphorylating aminoacyl-tRNA synthetases in *vivo*: HipA2 targets lysyl-tRNA synthetase (LysS), tryptophan-tRNA synthetase (TrpS), and GltX, while HipA1 phosphorylates TrpS and GltX (57). In both cases, the stringent response regulator SpoT (Rel) is required for *hipBA*1- and *hipBA*2-mediated persistence. A recent report confirmed that HipA2 only phosphorylates and inhibits TrpS, whereas LysS and GltX were not found to be phosphorylated (58). Nevertheless, both studies agreed that HipA2 induces the stringent response and persistence in parallel (57, 58).

We recently described a new family of minimal bacterial kinases, HipT, members of which exhibit sequence similarity with the C-terminal part of HipA but are encoded by
three-gene operons (Fig. 1B) (59). HipT of E. coli O127 is functionally similar to HipA and phosphorylates tryptophanyl-tRNA synthetase (TrpS/TrpRS) at two conserved serine residues, inactivating the enzyme. Likewise, ectopic production of hipT inhibits cell growth and translation and, consistently, stimulates production of (p)ppGpp (59). The gene immediately upstream of hipT encodes a small protein, HipS, that exhibits sequence similarity to the N-terminal part of the larger HipA kinase (Fig. 1B). Surprisingly, HipS neutralizes HipT in vivo, and therefore appears to function as the antitoxin of the hipBST module. The third component, HipB, encoded by the first gene of the hipBST operon (Fig. 1B), contains an HTH domain and is homologous to HipB of HipBA but does not counteract HipT kinase activity directly. Rather, this protein functionally appears to augment the ability of HipS to neutralize HipT (59). The structural and mechanistic details that set the bicistronic and tricistronic Hip kinase systems apart have not yet been elucidated. Finally, E. coli K-12 carries genes that encode the HipA homologue YjjJ in a monocistronic operon, thus lacking an adjacent antitoxin gene, which also inhibits cell growth upon induction (Fig. 1C) (60). Interestingly, YjjJ contains a HTH domain at its N terminus that may compensate for the lack of a DNA-binding antitoxin; however, how YjjJ kinase activity is controlled remains unknown.

The discovery of the hipBST tricistronic operons (59) as well as the observed diversity among Hip toxin homologues in various bacterial species inspired us to investigate the overall phylogeny of HipA-homologous proteins and their gene families in prokaryotic microorganisms. This led to the discovery of seven novel Hip kinase families, potential antitoxins with novel features such as HIRAN (HIP116 and RAD5 N-terminal) domains with predicted specificity for single-stranded or double-stranded DNA ends and a novel putative two-domain antitoxin family consisting of a HipS-like domain coupled to a HIRAN domain. We also find evidence that HipA-homologous kinases are present in Archaea. Together, these results delineate the structural and functional diversity of the family of HipA kinases and suggest directions for future experimental research.

RESULTS AND DISCUSSION

HipA-homologous kinases form a strongly supported, bifurcated phylogenetic tree. The phylogenetic analysis was initiated using HipA and YjjJ of E. coli K-12 and HipT of E. coli O127 as seed sequences using BLASTP and HMMSEARCH (see Materials and Methods for details). While this revealed a vast number of HipA-homologous kinases within the bacterial domain and a few in the archaeal domain, searches in the Eukarya domains did not disclose significant homologues (E<10^-5). To analyze the vast number of high-score homologues (E>10^-19) systematically, we repeated the search using individual bacterial and archaeal phyla as search spaces (see Fig. S1 in the supplemental material) and retrieved ~1,800 high-scoring Hip homologues. Tenacious curation, including manual inspection of putative neighboring antitoxin genes of each individual kinase gene, removal of incomplete genes, and exclusion of closely related kinases (5% sequence difference) reduced the number of included kinases to a final set of 1,239 sequences. The majority of these are from the phylum Proteobacteria (70%) while they are also frequently observed among Actinobacteria (13%), Firmicutes (5%), and Spirochaetes (2%) (see Table S1 in the supplemental material). Using these sequences, we generated a phylogenetic tree, called the “Hip Tree” (Fig. 2A). Fully annotated and bootstrapped versions of the Hip Tree are shown in Fig. S2A and S2B, while a full breakdown of the phylogeny of the kinases is given in Table S2A. The Hip Tree consists of 11 main clades (clades I to XI) all supported by high statistical confidence levels (Table 1). The tree is bifurcated with one branch containing main clades I through X, while the other branch consists of the diverse main clade XI (Fig. 2A). In Fig. 2A, the colors of the clades (shown as triangles) reflect kinases encoded by TA modules with identical genetic organization, as explained in the next section. In other words, each differently colored triangle in Fig. 2A reflects HipA-homologous kinases encoded by TA modules with a different genetic organization.

HipA-homologous kinases are encoded by 10 different genetic organizations. By careful investigation of the sequence data set in Table S1, we found that a
consistent and biological meaningful definition of “Hip kinase family” can be based on the genetic context encoding the kinases. Using this classification, we identified a total of 10 Hip kinase families encoded by TA modules with 10 different genetic organizations (Fig. 2B). The frequencies of the TA modules with different genetic organizations are given in Table S2B.

Using this classification, all HipA kinases encoded by hipBA modules with an upstream HipB HTH antitoxin gene cluster together in clades I, II, and IV to VIII. Experimentally characterized hipBA modules from *E. coli* K-12 and *C. crescentus* are all in clade I (Fig. S2A). Interestingly, clades VI and VII each contain a subclade that have a reversed gene order (i.e., hipAB) that is—when the gene order hipAB. The Hip Tree was visualized by iTOL (82). (B) Genetic organizations of the TA modules encoding the 1,239 HipA-homologous kinases. The various types of genetic organization were obtained by manual inspection of the genes upstream and downstream of the kinase genes listed in Table S1 in the supplemental material. The coloring of the Hip kinases in panel B follows the coloring of the clades in panel A. Putative antitoxins with helix-turn-helix (HTH) domains are colored light green. Stl/HTH, putative antitoxins containing HTH domain and a domain with structural similarity to the “polyamorous” repressor Stl encoded by *Staphylococcus aureus*; HipS, HipS-like; HIRAN, HIP116 Rad5p N-terminal domain.

**FIG 2** Phylogeny and genetic contexts of HipA-homologous kinases. (A) Simplified phylogenetic tree covering 1,239 HipA-homologous kinases (the “Hip Tree”) (see Fig. S2 in the supplemental material for details). The Hip Tree was divided into 11 main clades I to XI. The coloring of the Hip Tree reflects the genetic contexts that encode the kinases such that each color corresponds to a distinct kinase family encoded by a distinct type of TA module. All main clades are monophyletic except clade XI that consists of six different kinase families. Small blue triangles within the red triangles of main clades VI and VII symbolize subclades of kinases encoded by TA modules with a reversed gene order relative to hipBA—that is—with the gene order hipAB. The Hip Tree was visualized by iTOL (82). (B) Genetic organizations of the TA modules encoding the 1,239 HipA-homologous kinases. The various types of genetic organization were obtained by manual inspection of the genes upstream and downstream of the kinase genes listed in Table S1 in the supplemental material. The coloring of the Hip kinases in panel B follows the coloring of the clades in panel A. Putative antitoxins with helix-turn-helix (HTH) domains are colored light green. Stl/HTH, putative antitoxins containing HTH domain and a domain with structural similarity to the “polyamorous” repressor Stl encoded by *Staphylococcus aureus*; HipS, HipS-like; HIRAN, HIP116 Rad5p N-terminal domain.
but does not function as an antitoxin as mentioned above (20, 59). As described later, HipR-encoded *hipRF* and *hipRG* operons also contain a HTH domain but exhibits structural similarity with the Stl repressor.

**Conserved HipA kinases appear to contain functionally significant differences.** Clade I contains the “classical” bacterial HipA kinases with known cellular targets, HipA of *E. coli* K-12, HipA1, and HipA2 of *C. crescentus*, together with their close homologues. An alignment of representative sequences of subclades containing HipA, HipA1, and HipA2 kinases reveals, as expected, the four canonical core kinase motifs: the Gly-rich loop, the activation loop, the catalytic motif, and the Mg2⁺-binding motif (Fig. S3). The alignment also reveals a number of insertions (ω1 to ω6) and deletions (Δ1 to Δ3), also called “indels” in the two C. crescentus HipA (HipA1, HipA2) subclades relative to the *E. coli* K-12 HipA (HipA1, HipA2) subclade (Fig. S3). Figure 3A shows a schematic overview of HipA indicating the positions of the indels relative to the core kinase motifs in the primary sequences, while Fig. 3B and C show a mapping of the HipA1- and HipA2-specific indels as well as regions of high sequence divergence onto the structure of HipA of *E. coli* K-12. Interestingly, despite being distant in the primary sequence, the two largest deletions in *E. coli* HipA (Δ2 and Δ3) are adjacent in the tertiary structure (Fig. 3B and C) and close to the γ-phosphate of ATP. Additionally, both *C. crescentus* HipA kinases share a C-terminal region of high sequence divergence that maps to solvent-exposed residues of a surface helix (Fig. 3A), while HipA2 has an additional unique region of divergence (Fig. 3C). Studies of eukaryotic cyclin-dependent kinases have shown that the homologous region where Δ2, Δ3, and both regions of divergence are situated is involved in target binding (61), raising the possibility that the differences observed relate to differences in target specificity. Finally, we note that HipA2 has several regions (ω1′, ω2′, and ω3′) not present in HipA (Fig. 3A and Fig. S3) concentrated in the region that interacts with the very C terminus of HipB in the *E. coli* HipB structure (Fig. 3C), thus potentially affecting the mechanism by which the antitoxin interacts with and inhibits the cognate kinase.

**All HipT kinases lack the canonical N-terminal subdomain.** Main clade III consists exclusively of HipT kinases (Fig. 2). As shown before, the sequences of the HipT kinases of *E. coli* O127, *Haemophilus influenzae*, and *Tolumonas auensis* align colinearly with the

---

**TABLE 1** Overview of the clades of the Hip Tree consisting of 1,239 kinases.

<table>
<thead>
<tr>
<th>Clade</th>
<th>No. of kinase genes</th>
<th>TA gene organization</th>
<th>Experimentally analyzed TA modules</th>
<th>Cellular target(s) of toxin</th>
<th>GenBank ID</th>
<th>Reference(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>712</td>
<td><em>hipBA</em></td>
<td><em>hipBA</em> of <em>E. coli</em> K-12</td>
<td>GltX</td>
<td>NP_416024.1 (440)</td>
<td>47, 51, 52, 57, 58</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td><em>hipBA</em>, of <em>C. crescentus</em></td>
<td>Unknown or GltX and LysS</td>
<td>ACL93947.1 (423)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td><em>hipBA</em>, of <em>C. crescentus</em></td>
<td>LysS or GltX, LysS and TrpS*</td>
<td>ACL96286.1 (444)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td><em>hipBA</em>, of <em>C. crescentus</em></td>
<td>Unknown*</td>
<td>WP_010920611.1 (435)*</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td><em>hipBA</em>, of <em>S. oneidensis</em> MR-1</td>
<td>Unknown</td>
<td>AAN53784.1 (433)</td>
<td></td>
</tr>
<tr>
<td>II</td>
<td>73</td>
<td><em>hipBA</em></td>
<td>None</td>
<td>TrpS</td>
<td>WP_015879003.1 (342)</td>
<td>59</td>
</tr>
<tr>
<td>III</td>
<td>48</td>
<td><em>hipBA</em></td>
<td><em>hipBST</em> of <em>E. coli</em> O127</td>
<td>TrpS</td>
<td>NP_438824.1 (343)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td><em>H. influenzae</em> KW20</td>
<td>TrpS</td>
<td>CAS11333.1 (335)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td><em>T. auensis</em> DSM9187</td>
<td>TrpS</td>
<td></td>
<td></td>
</tr>
<tr>
<td>IV</td>
<td>14</td>
<td><em>hipBA</em></td>
<td>None</td>
<td>None</td>
<td></td>
<td></td>
</tr>
<tr>
<td>V</td>
<td>9</td>
<td><em>hipBA</em></td>
<td>None</td>
<td>None</td>
<td></td>
<td></td>
</tr>
<tr>
<td>VI</td>
<td>36</td>
<td><em>hipBA</em></td>
<td>None</td>
<td>None</td>
<td></td>
<td></td>
</tr>
<tr>
<td>VII</td>
<td>132</td>
<td><em>hipAB</em> and <em>hipAB</em></td>
<td>None</td>
<td>None</td>
<td></td>
<td></td>
</tr>
<tr>
<td>VIII</td>
<td>21</td>
<td><em>hipBA</em></td>
<td>None</td>
<td>None</td>
<td></td>
<td></td>
</tr>
<tr>
<td>IX</td>
<td>12</td>
<td><em>hipEB</em></td>
<td>None</td>
<td>None</td>
<td></td>
<td></td>
</tr>
<tr>
<td>X</td>
<td>101</td>
<td><em>hipHYjjj</em></td>
<td><em>hipHYjjj</em></td>
<td>None</td>
<td>P39410.1 (443)</td>
<td>60</td>
</tr>
<tr>
<td>XI</td>
<td>81</td>
<td><em>hipRF</em>, <em>hipRG</em>, <em>hipMP</em>, <em>hipL</em>, <em>hipN</em>, and <em>hipS</em></td>
<td>None</td>
<td>None</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Data compiled from Table S1 in the supplemental material.

*Gene organization refers to information retrieved from Table S1 and visualized schematically in Fig. 2B.

*GenBank identifiers (IDs) or accession numbers of the kinases are from Table S1 and their position in amino acids in the Hip Tree visualized in Fig. S2A and S2B in the supplemental material.

*hipBA* of *C. crescentus* previously had the accession number NP_421566.1 and was used in Table S1.

*The substrate of HipA1 is either unknown (58) or GltX plus LysS (57), while the substrate of HipA2 is either TrpS (58) or GltX, LysS plus TrpS (57).
C-terminal part of *E. coli* K-12 HipA (Fig. S4A) but lack the canonical ~100-amino-acid (aa) N-subdomain-1 that serves as a lid on top of the core kinase domain in HipA (59). Nevertheless, the four conserved Hip kinase motifs (Gly-rich loop, activation loop, catalytic motif, and Mg²⁺-binding motif) are conserved in the HipTs, as well as a serine adjacent to the Gly-rich loop that is subject to autophosphorylation in both HipT and HipA (46, 59). We showed previously that HipS, which is encoded immediately upstream of HipT, functions as the antitoxin of *hipBST* modules and that this protein exhibits sequence similarity to the N-terminal domain of HipA (59) (Fig. S4A and S4B). In other words, the “missing” N-subdomain-1 of HipT appears to be encoded immediately upstream and thus in a similar genomic location relative to the core kinase domain as in HipA, which suggests that the gene has been split (or merged) at some point during evolution. However, the functional and structural implications of this difference with respect to kinase activation and regulation are not yet understood.

**YjjJ kinases contain an N-terminal HTH domain.** YjjJ of *E. coli* K-12 is a HipA homologue encoded by a monocistronic operon and thus “lacking” a closely linked antitoxin or DNA-binding HipB-like gene. Instead, YjjJ has a ~100-aa N-terminal extension containing a helix-turn-helix (HTH) domain (residues 15 to 34) that potentially could function as a cis-acting antitoxin, a property previously found with other type II modules (62). To maintain a uniform nomenclature, we propose here to rename YjjJ kinases HipH (H for HTH domain) as these constitute the monophyletic clade X in the Hip Tree (Fig. 2A). To avoid alignment of nonhomologous domains, we chose to align subclade HipH$_{E. coli}$ K-12 with HipT$_{E. coli}$ O127 because HipT lacks the ~100-aa N-terminal domain present in HipB but has all the canonical kinase domains. As seen from Fig. 4,
HipH and HipT kinases align colinearly with respect to the four conserved kinase motifs, while the N terminus of HipH forms a separate HTH domain. Moreover, HipH kinases contain two conserved serine residues near the Gly-rich loop, raising the possibility that they are regulated through autophosphorylation in a way similar to HipT (59).

**Main clade XI consists of kinases belonging to six different families.** The kinases in main clade XI of the Hip Tree are encoded by six different genetic contexts and define six novel Hip kinase families (Fig. 2A and B). All six subclades are separated by high bootstrap values, supporting that their separation into novel kinase families is phylogenetically justified (Fig. S5). The six families encompass four families of short kinases (HipG, HipM, HipI, and HipJ), one longer variant (HipF) similar in size to HipA and a “long” kinase family (HipL) (Fig. 2B). In the following, we describe these six new types of kinase-encoding TA modules.

**The HipG, HipM, HipI, and HipJ families contain the core kinase domain.** The HipG, HipI, HipJ, and HipM kinases are all phylogenetically closely related and are, on average, even smaller than the HipT kinases (279 to 321 aa versus 291 to 346 aa, respectively) (Table S1 and Fig. S5). A sequence alignment of representatives of these families confirms that they contain the four conserved core Hip sequence motifs are thus likely active kinases (Fig. S6). However, like HipT, these kinases lack the N-subdomain-1 present in HipA, HipE, and HipF and consequently, the Gly-rich loop is located close to the N termini (Fig. S6). The kinases in the HipG group do not have a conserved serine or threonine adjacent to the Gly-rich loop, suggesting they are probably not regulated by autophosphorylation in a way similar to HipA. In contrast, members of the other families (HipM, HipI, and HipJ) mostly have either Ser or Thr near the Gly-rich loop, suggesting on the other hand that they may be regulated in this way. We note that HipI kinases constitute four separate subclades of the phylogenetic reconstruction of main clade XI (Fig. 2A). Three of the subclades consists of kinases from *Actinobacteria* and *Firmicutes* while the fourth consists of a mix of kinases from *Proteobacteria* and Cyanobacteria (Fig. S5).

**HipF and HipG kinases have putative antitoxins related to polyamorous repressor Stl.** HipF and HipG kinases are encoded in operons with putative antitoxins that we call HipR (Fig. 2B). Interestingly, Phyre2 (63) predicts HipR to be structurally related to the two-domain transcriptional repressor Stl encoded by *Staphylococcus aureus* superantigen-carrying pathogenicity islands (SaPl). Stl has an HTH domain and maintains integration of SaPl elements in the bacterial chromosome by transcriptionally repressing genes essential for element excision and replication (64). SaPl excision and replication are induced by invading phages via specific interaction between Stl and nonessential phage proteins (65–69). These observations allow us to hypothesize that phage proteins could potentially activate HipF and/or HipG kinases during infection via interaction with HipR. Under this model, activation of the kinases induces abortive phage infection by inhibition of translation via phosphorylation of an essential component of the protein synthesis apparatus, thus eliciting phage resistance. This possibility can now be tested experimentally.

**Putative HipN antitoxins consist of a HipS-like and a HIRAN domain.** The genes encoding HipM, HipI, and HipJ kinases all have downstream genes coding for putative antitoxins (Fig. 2B). The putative antitoxin HipN encoded by *hipIN* is a two-domain protein consisting of an N-terminal HipS-like domain (Fig. S7A) and a C-terminal HIRAN domain (Fig. S8A). The HipS-like domain of HipN may function to neutralize its cognate HipI kinase as is the case of HipS encoded by *hipBST* (59). A possible function of the HIRAN domain of HipN is discussed below.

**A second family of putative HipS antitoxins.** *hipJS* modules encode a HipJ kinase and a putative antitoxin HipS that exhibits similarity to HipS of *hipBST* (Fig. S7A). Thus, similarly to HipT (59), HipJ may be neutralized by its cognate HipS. As noted above, HipS and HipS-like domains exhibit sequence similarity with the ∼100-aa N-subdomain-1 of HipA, and this domain was therefore included in the alignment of the HipS antitoxins and HipS-like domains (Fig. S7A). The phylogenetic tree based on the HipS and HipS-like sequences is bifurcated, with HipS encoded by *hipBST* and *hipJS* branch together with N-subdomain-1 of HipA, while the HipS-like sequences encoded by *hipl* and *hipN* generate a
FIG 4 Alignment of HipH and HipT kinases. Sequence alignment of subclades containing HipT of *E. coli* O127 and HipH of *E. coli* K-12. Deletions (Δ) and insertions (ω) relative to the HipH* E. coli* K-12 subclade are indicated, as well as the four conserved kinase motifs (Gly-rich loop, activation loop, catalytic motif, and Mg\(^{2+}\)-binding motif). HTH domains in the N terminus of HipH kinases are boxed in red.
distinct second branch (Fig. S7B). HipJ kinases are closely related to HipI kinases (Fig. S5), and it is tempting to speculate that hipJS modules evolved from hipIN modules by deletion of the HIRAN domain of HipN antitoxins (Fig. 2B).

**HipL kinases contain both HipS-like and HIRAN domains.** HipL kinases, encoded by monocistronic operons, form a single subclade of main clade XI that is further divided into two subclades (Fig. S5). The two subclades consist of HipL kinases from Gram-positive and Gram-negative bacteria. Their sequences are clearly distinct, with many subclade-specific insertions and deletions in the kinase core domain (Fig. S8B). Alignment of HipL and HipA reveals that HipL maintains the four, conserved core kinase motifs and has a large C-terminal extension of ~200 aa of which the ~100 aa at the extreme C terminus are annotated at GenBank as a HIRAN domain (Fig. S8C). Like the small kinases, HipL kinases lack the N-subdomain-1 of ~100 aa relative to the core kinase domain of HipA (Fig. S8C). Unexpectedly, however, the ~100-aa domain located between the core kinase domain and the C-terminal HIRAN domain exhibits sequence similarity with HipS (Fig. S8D). The N-terminal sequences of HipA also align well with the HipS-like domain of HipL (Fig. S8E), consistent with the fact that HipS exhibits sequence similarity with the N-subdomain-1 of HipA (Fig. S4B). In summary, HipL kinases are three-domain proteins consisting of an N-terminal kinase core domain, a HipS-like domain, and a C-terminal HIRAN domain (Fig. 2B and Fig. S8C). We note the possibility that the HipS-like domain may be involved in regulating the kinase activity of HipL, reminiscent of how HipS regulates the kinase activity of HipT (59).

The HIRAN domains of HipL and HipN may bind DNA. The presence of HIRAN domains in both HipL kinase and the putative antitoxins HipN is interesting, not least because HIRAN domains can bind DNA. Moreover, the HIRAN domains of HipL kinases and HipN antitoxins are in both cases joined with a HipS-like domain (Fig. 2B). HIRAN domains have previously been identified in eukaryotic multidomain DNA repair proteins (70). Experimentally analyzed HIRAN domains bind single-stranded or double-stranded DNA ends (71, 72). Structural studies of human helicase-like transcription factor (HLTF), a DNA helicase implicated in remodeling of replication forks, including fork regression and restart (73), revealed the residues required for DNA binding (72) (Fig. 5A and B). Due to a high sequence divergence, we decided to split the HIRAN domains of the HipN homologues based on their phylogenetic origin (Fig. 5C and Fig. S8A). Most of the HIRAN domains retained the majority of the sequence motifs that interact with DNA, with the exception of HipN from *Actinobacteria*. A study of the HIRAN domain of human HLTF showed that Phe142 (NAE) is required for binding to duplex DNA because it stacks with nucleobases of the other strand (72) (Fig. 5B). Importantly, most of the bacterial HIRAN domains lack a conserved Phe at this position (Fig. 5C), raising the possibility that the HIRAN domains may be involved in triggering the kinase activity of HipL, reminiscent of how HipS regulates the kinase activity of HipT (59).

**A putative antitoxin with a s-resolvase-like domain.** A final variation among the Hip kinases is found for the hipMP modules, which encode a HipM kinase that is short and closely related to HipG kinases (Fig. 2A) and a putative HipP antitoxin that exhibits similarity to γσ-resolvases (Fig. 2B). Bacterial γσ-resolvases are transposon-encoded enzymes that catalyze recombination within a complex nucleoprotein structure during site-specific DNA recombination and thus have the capability to bind DNA (76). Interestingly, γσ-resolvases are structurally related to 5’–3’ exonucleases active on RNA, which may also provide clues to the role of this domain in the context of Hip kinases (77). By analogy with other hipBA modules (28, 78), we therefore postulate that
HipP functions as antitoxin in the hipMP systems and may bind DNA to autoregulate transcription as seen for many type II TA systems.

**Conclusion.** In addition to the three known kinase families HipA, HipT, and YjjJ/HipH, we discover here seven novel Hip kinase families encoded by different genetic contexts. Kinases of one family, HipE, are encoded by TA operons with a reversed gene order relative to hipBA, while kinases of families HipF and HipG have putative Stl-homologous antitoxins that may be regulated by proteins encoded by attacking phages. Kinases of one family, HipJ, are associated with HipS domain putative antitoxins that, similar to HipS of hipBST, may interact with and neutralize their cognate kinase toxins. Kinases of yet another family, HipI, are associated with putative antitoxins that consist of a HipS-like domain and a HIRAN DNA-binding domain. Finally, HipL kinases, encoded by monocistronic operons, consist of an N-terminal core kinase domain, an internal HipS-like domain, and a C-terminal HIRAN DNA-binding domain. Our analysis presented here builds a foundation for the future experimental analysis of HipA-homologous kinases.

**MATERIALS AND METHODS**

**Data sampling.** As of December 2019, HipA and HipH/YjjJ of *E. coli* K-12 and HipT of *E. coli* O127 were used as seeds in BLASTP searches at https://blast.ncbi.nlm.nih.gov/, using the bacterial phyla shown in Fig. S1 as search spaces. HMMSEARCH at www.ebi.ac.uk (79) was used to expand poorly populated clades. In total, ≈1,800 Hip sequences were retrieved (E > 10^-20) and curated manually such that every Hip kinase sequence retained satisfied the following criteria. (i) The Hip kinase gene should encode a full-length kinase with the four canonical kinase motifs (Gly-rich loop, activation loop, catalytic motif, and Mg^2+ binding domain), as determined from a multiple-sequence alignment (MSA). (ii) Kept kinases...
should be encoded by a gene with an adjacent upstream or downstream putative antitoxin gene. (iii) In general, the adjacent, putative antitoxin gene should encode a DNA-binding protein (although this criterion is not satisfied by HipJ proteins). In Table S1 in the supplemental material, kept kinases are less than 95% identical to any other kept kinase. By this scrutiny, the initial gene set was reduced to 1,239 Hip gene modules (Table S1 and Fig. 2).

**Toxin-antitoxin module gene organization.** The gene organizations shown in Fig. 1 were deduced by manual inspection of genes neighboring the Hip kinase genes of Table S1.

**Sequence alignments and phylogenetic tree reconstruction.** Sequence alignments were generated by Clustal Omega (80) at www.ebi.ac.uk and imported into Jalview (81). Protein sequence alignments in Jalview 2.11.0 were exported as vector files (EPS or SVG formats) and imported into Adobe Illustrator CS6, annotated, and saved in PDF format for publication. Phylogenetic trees were visualized using iTOL (82). Reconstruction of phylogenetic trees was accomplished using IQ-TREE that uses the maximum likelihood approach and Ultrafast bootstrapping via the CIPRES module in Genious Prime (83–85). The kinase sequence alignments used in the reconstruction of the three phylogenetic trees that we present had the following characteristics. (i) The sequence alignment of the 1,239 kinases (Fig. S2) has 1,556 columns, 1,511 distinct patterns, 1,172 parsimony-informative sites, 172 singleton sites, and 211 constant sites. (ii) The sequence alignment of the 81 sequences of main clade XI (Fig. S5) has 785 columns, 761 distinct patterns, 650 parsimony-informative sites, 58 singleton sites, and 77 constant sites. (iii) The sequence alignment of the 112 HipS and HipS-like sequences (Fig. S7B) has 170 columns, 170 distinct patterns, 146 parsimony-informative sites, 16 singleton sites, and 8 constant sites. The alignment of the HipS and HipS-like sequences has fewer singleton and constant sites, thus explaining the low bootstrap values (Fig. S7B).

Structure similarity searches were done using Phyre2 (63) (http://www.sbg.bio.ic.ac.uk/phyre2/) and mapping of deletions and insertions on existing structures using PyMOL. HTH motifs were identified by two different algorithms, EMBOSS (86) and HELIX-TURN-HELIX MOTIF PREDICTION (87).

**SUPPLEMENTAL MATERIAL**

Supplemental material is available online only.

**FIG S1**, PDF file, 2.5 MB.
**FIG S2**, PDF file, 2.1 MB.
**FIG S3**, JPG file, 2.5 MB.
**FIG S4**, PDF file, 2.7 MB.
**FIG S5**, PDF file, 0.3 MB.
**FIG S6**, JPG file, 2.8 MB.
**FIG S7**, JPG file, 2.6 MB.
**FIG S8**, JPG file, 2.8 MB.
**TABLE S1**, TXT file, 0.7 MB.
**TABLE S2**, DOCX file, 0.02 MB.

**ACKNOWLEDGMENTS**

We thank Boris Macek (Proteome Center Tübingen, University of Tübingen, Germany) and Yong E. Zhang (Department of Biology, University of Copenhagen, Denmark) for critical comments on the manuscript.

This work was supported by a Novo Nordisk Foundation Ascending Investigator grant to D.E.B. (grant NNF18OC0030646), a Center-of-Excellence grant from the Danish Natural Research Foundation to K.G. (DNRF120), and a personal Laureate Research Grant from the Novo Nordisk Foundation to K.G.

**REFERENCES**


