NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH Public Access
Author Manuscript
Science. Author manuscript; available in PMC 2009 January 11.

Published in final edited form as:
Science. 2008 July 11; 321(5886): 256–259. doi:10.1126/science.1160033.

Genetic Determinants of Self Identity and Social Recognition in Bacteria*

Karine A. Gibbs[1,3], **Mark L. Urbanowski**[2,3], and **E. Peter Greenberg**[1,†]

1*Department of Microbiology, School of Medicine, University of Washington, Seattle, WA 98195*

2*Department of Microbiology, Carver College of Medicine, University of Iowa, Iowa City, IA 52242*

Abstract

The bacterium *Proteus mirabilis* is capable of movement on solid surfaces by a type of motility called swarming. Boundaries form between swarming colonies of different *P. mirabilis* strains but not between colonies of a single strain. A fundamental requirement for boundary formation is the ability to discriminate between self and non-self. We have isolated mutants that form boundaries with their parent. The mutations map within a six-gene locus that we term *ids* for identification of self. Five of the genes in the *ids* locus are required for recognition of the parent strain as self. Three of the *ids* genes are interchangeable between strains and two encode specific molecular identifiers.

About 60 years ago, different clinical isolates of the swarming bacterial species *Proteus mirabilis* were shown to form visually apparent boundaries between colonies growing on agar (1). By contrast, swarms of a single strain merge with each other (2,Fig. 1A). This phenomenon is still used in diagnostic laboratories to type clinical isolates of *P. mirabilis* (3). Many clinical isolates of *P. mirabilis* secrete proteins called proticines that kill sensitive strains. An individual strain of *P. mirabilis* can be identified by a combination of the proticine it produces and the proticines to which it is sensitive (4,5). Boundaries form between swarms of strains differing in proticine production and sensitivity. However some strains do not produce any proticines but still form boundaries, even with other non-proticine producing strains. Thus proticine production and sensitivity do not explain boundary formation. We sought to identify self versus non-self discrimination factors required for boundary formation by screening for and isolating mutants that recognize their parent as different from self.

We chose *P. mirabilis* strain BB2000 as a model because it is genetically tractable (6). We used an agar plate assay to screen 3600 BB2000 mutants, generated by random transposon mutagenesis, in a format where each mutant swarm had two, three or four adjacent neighbors (2). We discovered a single mutant that formed a boundary with every adjacent mutant, and we named the mutant phenotype "identification of self (Ids)" because mutant and parent swarms did not merge with each other. To show that the transposon insertion was responsible for the phenotype, we crossed the insertion in the Ids transposon mutant into the BB2000 parent by homologous recombination and isolated four recombinants, all of which formed boundaries with the parent but not with each other (Fig. S1).

*This manuscript has been accepted for publication in *Science*. This version has not undergone final editing. Please refer to the complete version of record at http://www.sciencemag.org/. Their manuscript may not be reproduced or used in any manner that does not fall within the fair use provisions of the Copyright Act without the prior, written permission of AAAS.

†Corresponding author: E. Peter Greenberg Department of Microbiology School of Medicine University of Washington 1705 NE Pacific Street Seattle, WA 98195−7242 Phone: 206.616.2881 FAX: 206.616.2983 Email: epgreen@u.washington.edu.
3These authors contributed equally to this work.

Supporting Online Material www.sciencemag.org Materials and Methods Figs. S1 to S5 References

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Boundaries between strain BB2000 and the independent isolate HI4320 contained individual cells of both strains at a low density as well as round bodies and debris. Cells of BB2000 and HI4320 made contact with each other within the boundary, but we did not observe cells that penetrated the opposite swarm (Fig. 1B). In boundaries between swarms of the Ids transposon mutant and the BB2000 parent we also observed a low density of cells, but round bodies and debris were not evident. Cells from the BB2000 parent swarm appeared to traverse the boundary and penetrate the Ids transposon mutant swarm (Fig. S2). During the merger of two swarms of strain BB2000, cells from each swarm penetrated the opposite swarm without apparent hindrance (Fig. 1C).

By using the sequenced genome of strain HI4320 (7), we found that the *ids* mutation mapped to codon 1030 of a 1033-codon open reading frame occurring between base pairs 3282912 to 3286013 and residing in a cluster of six genes (Fig. 2A). A homologous cluster was found by sequencing the parent strain BB2000 (2). We refer to the six-gene cluster as *idsABCDEF* for *id*entification of *s*elf. We constructed an *idsA-F* deletion mutant of strain BB2000 and found that boundaries formed between swarms of the *idsA-F* deletion mutant and the BB2000 parent, but not between the deletion mutant and the Ids transposon mutant (Fig. 1D). Complementation of the *idsA-F* deletion mutation with an *idsA-F* expression vector (which included the 800-bp region directly upstream of *idsA*) resulted in a transformant that merged with the BB2000 parent but formed boundaries with the deletion mutant (Fig. 1E). The complementation analysis confirmed that the *idsA-F* locus encodes self-recognition factors.

To assess individual *ids* gene functions, we introduced plasmids containing *idsA-F* gene clusters in which individual genes were disrupted into the *idsA-F* deletion mutant (2). We then tested all of the *ids*-plasmid-carrying strains to determine whether they merged with each other or formed boundaries on swarm plates (Fig. 2B). We classified the constructs into recognition groups that were comprised of strains whose swarms merged with each other but not with swarms of strains in different recognition groups (Fig. 2C). An *idsA*-deficient strain merged with swarms of wildtype BB2000 but formed boundaries with the *idsA-F* deletion mutant (Fig. 2B-C). In contrast, *idsB*, *idsC*, *idsD* or *idsE*-deficient strains merged with the *idsA-F* deletion mutant but formed boundaries with wildtype BB2000. The *idsF*-deficient mutant likewise formed boundaries with wildtype BB2000 but had the additional property of forming boundaries with the *idsA-F* deletion mutant and in fact, swarms of the *idsF*-deficient mutant formed boundaries with swarms of any construct but itself. We conclude that *idsA* is not required for recognition of the BB2000 parent as self, but *idsB*, *idsC*, *idsD*, *idsE* and *idsF* are required for self-recognition. The *idsF* gene appears to encode a recognition factor distinct in function from *idsB*, *idsC*, *idsD* or *idsE*-encoded factors as indicated by the fact that *idsF* mutants merged only with themselves.

To further investigate the function of the *ids* genes in self-recognition, DNA containing either the complete BB2000 *idsA-F* gene cluster or a combination of disruptions in the BB2000 *idsA-F* gene cluster were introduced into wildtype HI4320 by conjugation to create transgenic diploids (2). The diploid HI4320 strains partitioned into those that merged with wildtype HI4320 or those that formed boundaries with wildtype HI4320 (Fig. 2D and Fig. S3). A diploid HI4320 strain carrying the complete BB2000 *idsA-F* gene cluster formed boundaries with wildtype HI4320, but merged with swarms of diploid HI4320 strains carrying the BB2000 *idsA-F* gene cluster with disruptions of *idsA*, *idsB, idsC* or *idsF* (Fig. 2D). In contrast, swarms of diploid HI4320 strains carrying the BB2000 *idsA-F* gene cluster with disruptions in either *idsD* or *idsE* merged with wildtype HI4320 (Fig. 2D). Therefore *idsB*, *idsC* and *idsF* encode essential self-recognition functions and the *idsB*, *idsC* and *idsF* alleles can be complemented by alleles from a different strain. However, *idsD* and *idsE* are essential for self-recognition and appear to encode identity determinants.

NIH-PA Author Manuscript NIH-PA Author Manuscript NIH-PA Author Manuscript

To confirm that *idsD* and *idsE* encode identity determinants, DNA containing the HI4320 *idsA-F* gene cluster with gene disruptions in *idsD* and separately in *idsEF* were conjugated into wildtype BB2000 (2). Swarms of both the *idsD*-deficient and *idsEF*-deficient diploid BB2000 strains merged with wildtype BB2000 but formed boundaries with a diploid BB2000 strain carrying the complete HI4320 *idsA-F* gene cluster (Fig. 2D). Thus *idsD* and *idsE* encode identity determinants, which we refer to as molecular identifiers.

We note that a diploid BB2000 strain carrying the complete HI4320 *idsA-F* gene cluster formed boundaries with all other strains including the diploid HI4320 strain carrying the complete BB2000 *idsA-F* gene cluster (Fig. 2D). Therefore the *idsA-F* gene cluster is probably not the sole determinant of boundary formation between different strains. Consistent with the presence of additional unidentified determinants, boundaries formed even in situations where one of the swarming strains did not carry any of the *idsA-F* genes (i.e. the *idsA-F* deletion mutant).

The *idsA* and *idsB* genes encode polypeptides with significant sequence similarity to the conserved bacterial proteins Hcp and VgrG, respectively. Recently, *hcp* and *vgrG* were shown to form the first two genes in the type VI protein secretion system of *Vibrio cholerae* (8) and both *hcp* and *vgrG* homologs occur in multiple copies in many bacterial species including *P. mirabilis* (7,8,9). We have included *idsA* as part of the *ids* cluster even though it is not required for self-recognition, because it is linked to *idsB* homologs in other bacteria and because it is possible that another *hcp* homolog may be recruited to replace *idsA* in *idsA*-deficient *P. mirabilis* strains. The *idsC*, *idsD* and *idsE* gene products do not show significant similarity to other known polypeptides. The *idsF* gene encodes a conserved hypothetical bacterial protein.

We sequenced the *ids* loci from five additional isolates of *P. mirabilis*: CW677, CW977, G151, I5/5 and S4/3 (2). Swarms of the five strains formed boundaries with BB2000, HI4320, and each other. All strains had the six-gene *ids* locus, except strain CW677, which had a seven-gene *ids* locus that contained an additional gene with similarity to *idsE* (Fig. S4). In all strains, *idsA*, *idsB* and *idsC* were identical in length, and each polypeptide encoded by *idsA*, *idsB*, *idsC* or *idsF* had over 96% identity with its homologs from the other strains (Fig. 3A). Both IdsD and IdsE could be separated into two distinct subfamilies with 30% pair-wise identity. Within a single IdsD or IdsE subfamily, there was 97−99% pair-wise identity across the majority of the sequence. However within a subfamily, there was a C-terminal region in IdsD with only 72−84% pair-wise identity and a similar region of only 32−80% pair-wise identity in IdsE (between amino acids 80 and 169). The variable regions of *idsD* and *idsE* are reminiscent of alleles encoding antigenic variation in some bacterial pathogens (10).

The DNA immediately downstream of the *idsA-F* locus in strain BB2000 contains a gene coding for a polypeptide with sequence similarity to IdsF and two genes coding for polypeptides with similarity to IdsE (Fig. S4). We do not know if there are additional IdsE or F family members coded in the BB2000 genome, but the sequenced HI4320 genome contains a six-gene repeat between base pairs 84801 and 91381 coding for polypeptides with similarity to IdsE (7, Fig. S5). It is possible that the putative IdsE homologs could act as additional molecular identifiers.

We have not yet succeeded in detecting the products of any of the *ids* genes in *P. mirabilis* cells, and so we do not know their cellular locations or how they might function to allow swarms to discriminate themselves from other encroaching swarms. It is unlikely that this is a toxin-antitoxin system because we do not see evidence of dead cells in the boundaries between the Ids transposon mutant and its parent (Fig. S2) and because the *idsA-F* deletion mutant and the BB2000 parent grew equally well in mixed cultures. When inoculated at a 1:1 ratio, the ratio of the parent and the *idsA-F* deletion mutant in stationary phase remained 1:1. Instead, our data are consistent with a model for self-recognition in which *idsD* and *idsE* encode specific

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

molecular identifiers of self. The *idsB, idsC* and *idsF* products are devices necessary for self-non-self recognition and the *idsF* product has a function distinct from those of *idsB* and *idsC* (Fig. 3B).

Self-recognition may play a role in maintaining clonal *Proteus* infections (11). It also seems likely that other species of bacteria have genes encoding self-recognition. In fact, there is a report of swarm boundary formation between strains of the opportunistic pathogen *Pseudomonas aeruginosa* (12). The *P. mirabilis* genetic model of swarm identity provides a simplified system to further examine the molecular mechanisms of self-non-self recognition.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

References and Notes

1. Dienes L. Proceedings of the Society for Experimental Biology and Medicine 1946;63:265.
2. Materials and methods are available as supporting material on *Science* Online.
3. Sabbuba NA, Mahenthiralingam E, Stickler DJ. J Clin Microbiol 2003;41:4961. [PubMed: 14605124]
4. Senior BW. J Gen Microbiol 1977;102:235. [PubMed: 925679]
5. Senior BW. J Med Microbiol 1983;16:323. [PubMed: 6348290]
6. Belas R, Erskine D, Flaherty D. J Bacteriol 1991;173:6289. [PubMed: 1655704]
7. Pearson MM, et al. J Bacteriol 2008;190:4027. [PubMed: 18375554]
8. Pukatzki S, et al. Proc Natl Acad Sci U S A 2006;103:1528. [PubMed: 16432199]
9. Mougous JD, et al. Science 2006;312:1526. [PubMed: 16763151]
10. van der Woude MW, Baumler AJ. Clin. Microbiol. Rev 2004;17:581. [PubMed: 15258095]
11. Sabbuba NA, et al. J Urol 2004;171:1925. [PubMed: 15076313]
12. Munson EL, Pfaller MA, Doern GV. J Clin Microbiol 2002;40:4285. [PubMed: 12409411]
13. We thank Bernard Senior for his generous sharing of many *P. mirabilis* strains and helpful comments on the swarm assay. We also thank Robert Belas for providing strain BB2000, Harry Mobley for providing strain HI4320, and Melissa Visalli for providing strain G151. We thank Sudha Chugani, Breck Duerkop and Amy Schaefer for thoughtful scientific discussions and the W. M. Keck Foundation for support. K.A.G. was supported by training grant AI55396 from the National Institutes of Health. The sequences of the *ids* loci and flanking regions from strains BB2000, CW677, CW977, G151, I5/5 and S4/3 were deposited at GenBank, and the accession numbers are EU635876, EU635877, EU635878, EU635879, EU635880 and EU635881, respectively.

NIH-PA Author Manuscript
NIH-PA Author Manuscript
NIH-PA Author Manuscript

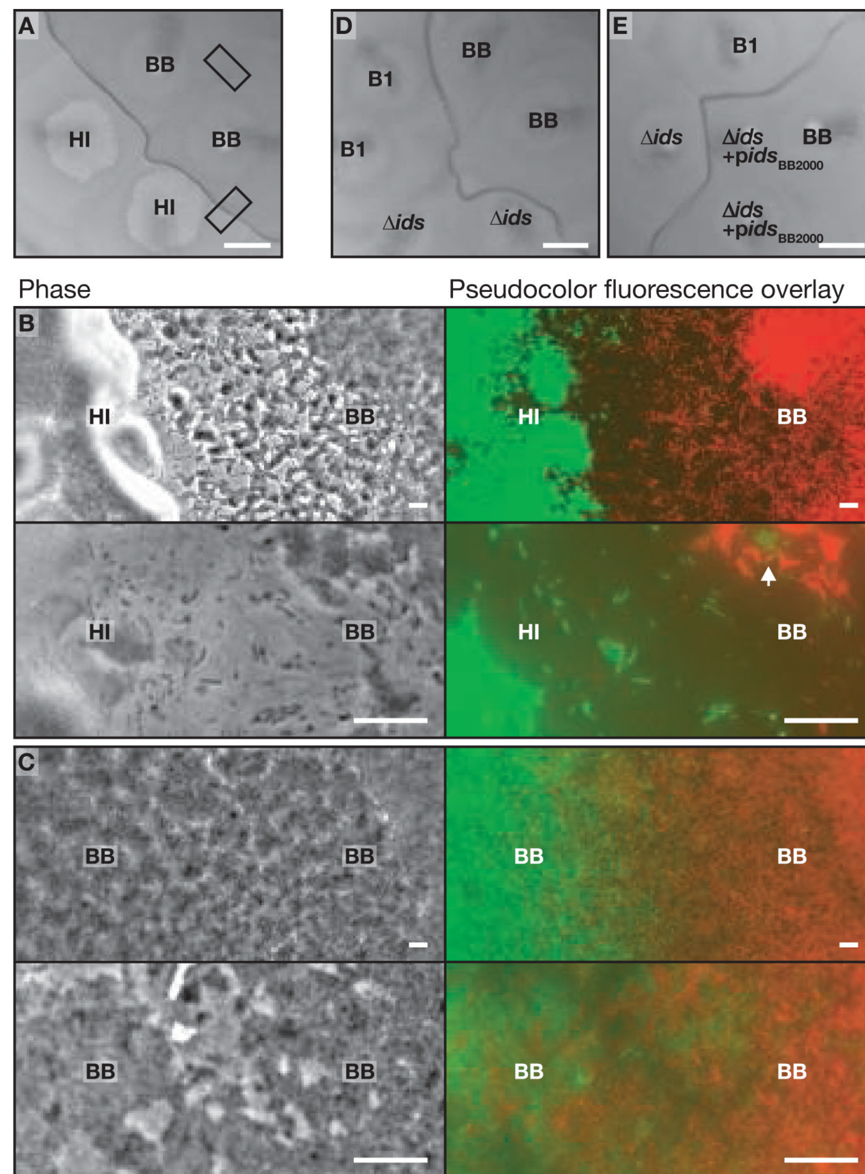

Figure 1. Images of swarm boundaries between different *P. mirabilis* strains

(A) Section of an agar plate with swarms of *P. mirabilis* strains HI4320 (HI) and BB2000 (BB). The boxes indicate intersections visualized in panels B and C. (B) Microscopy showing (top) the boundary between a GFP-labeled HI swarm (green) and a dsRed-labeled BB swarm (red), and (bottom) a higher magnification of the boundary. The arrow indicates HI cells among BB cells in the boundary. (C) Microscopy showing (top) the merger of two BB swarms, and (bottom) a higher magnification of the merger. The left BB swarm (green) expressed GFP and the right BB swarm (red) expressed dsRed (2). (D) Section of an agar plate with swarms of BB, the Ids transposon mutant (B1) and the *idsA-F* deletion mutant (*Δids*). (E) Section of an agar plate with swarms of B1, *Δids*, BB, and two swarms of the *idsA-F* deletion mutant carrying an *idsA-F* expression vector (*Δids*+p*ids*$_{BB2000}$). The bars are 1 cm (A, D and E) and 10 μm (B and C).

NIH-PA Author Manuscript
NIH-PA Author Manuscript
NIH-PA Author Manuscript

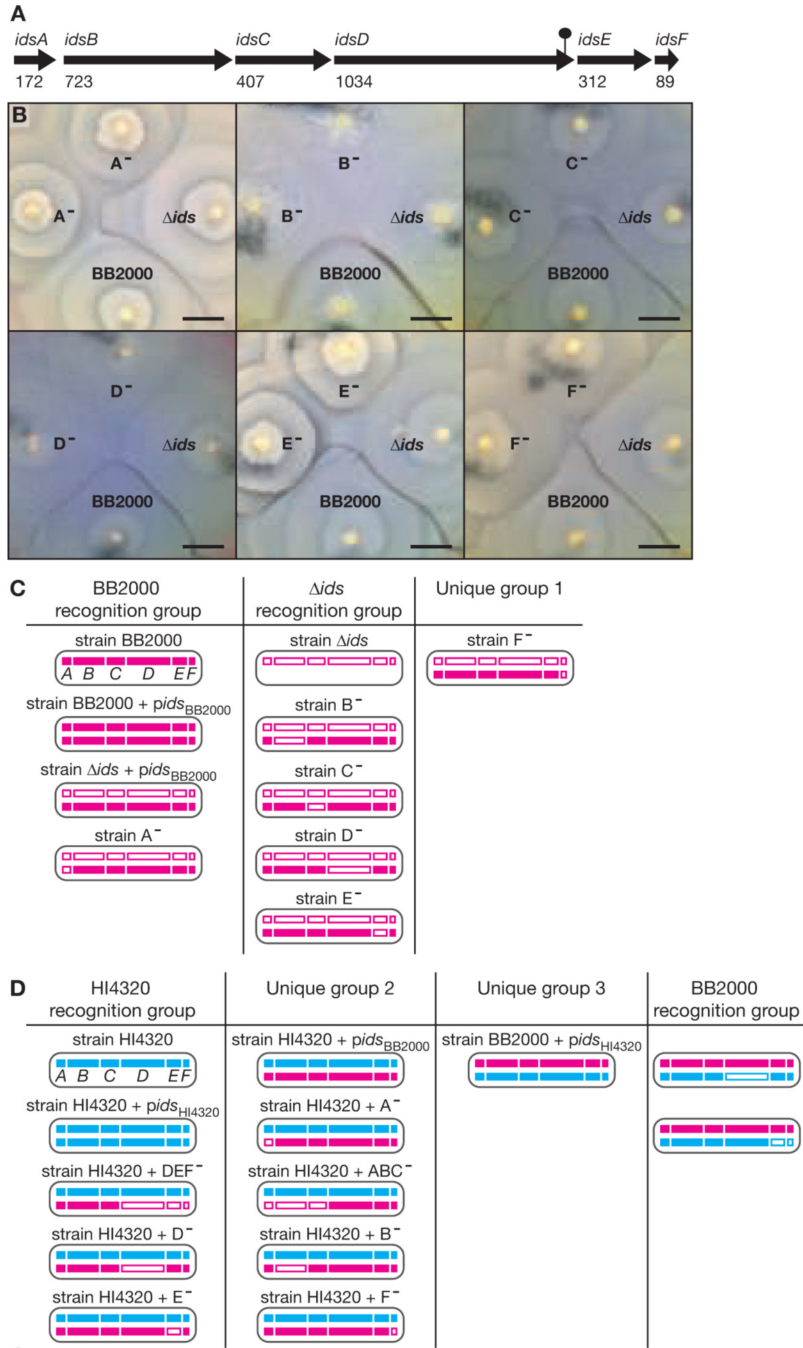

Figure 2. Genetic analysis of the *ids* gene cluster

(A) The length of the encoded polypeptides in the six-gene *ids* cluster of strain BB2000 is shown underneath each gene. The lollipop demarks the site of the transposon insertion in the Ids transposon mutant. (B) Sections of agar plates with the BB2000 parent, the *idsA-F* deletion mutant (Δ*ids*) and the *idsA-F* deletion mutant carrying a plasmid-borne BB2000 *idsA-F* gene cluster with single-gene disruptions in *idsA* (A$^-$), *idsB* (B$^-$), *idsC* (C$^-$), *idsD* (D$^-$), *idsE* (E$^-$) or *idsF* (F$^-$). The bars are 1 cm. (C) Recognition groups of strains constructed in the BB2000 and *Δids* backgrounds. A subset of the strains is shown in panel B. (D) Recognition groups of transgenic diploid derivatives of wildtype HI4320 and wildtype BB2000. Agar plates with swarms of representative strains are shown in Figure S3. For panels C and D, a cell of each

strain is represented by an oval in which the chromosomal and plasmid-borne *ids* gene clusters are on the top and bottom, respectively. A white box denotes a gene disruption. The BB2000 and HI4320 *ids* gene clusters are in red and blue, respectively. Each strain was tested against every other strain to determine the recognition groups (2). Each recognition group was comprised of strains whose swarms merged; boundaries formed between swarms of strains in different recognition groups.

NIH-PA Author Manuscript
NIH-PA Author Manuscript
NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

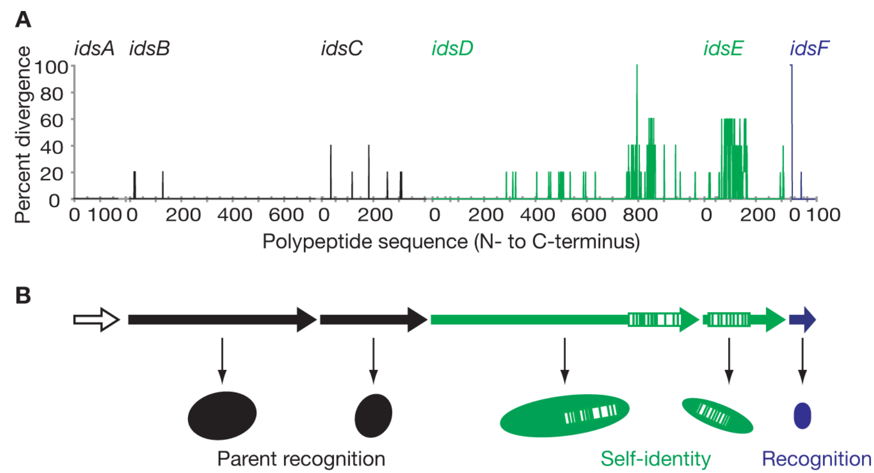

Figure 3. Organization and model of the *ids* gene cluster

(A) A plot showing the percent divergence of the encoded polypeptides in the *idsABCDEF* gene cluster amongst *P. mirabilis* strains BB2000, HI4320, CW977, G151 and S4/3. (B) A model for self-recognition. The *idsB*, *idsC*, *idsD*, *idsE* and *idsF* genes are required for recognition of the BB2000 parent as self. IdsA is not required for self-recognition. IdsF has a function in self-non-self recognition that is distinct from that of IdsB and IdsC. The *idsD* and *idsE* genes encode specific molecular identifiers of self.