Supporting Information

Mirouze et al. 10.1073/pnas.1219868110

SI Results and Discussion

Possible Significance of the Peculiar Evolution of DprA from Streptococci with ComRS-Regulated Competence. When comparing the rate of evolution of the sterile α-motif (SAM) domain to that of the extended Rossmann fold (eRF) domain, two behaviors were observed. The SAM domain from ComE-dependent competence streptococci, Lactobacillaceae, and *Leuconostoc* presented an accelerated evolution compared with the eRF domain (Fig. 5*C*). In constrast, the ComRS-dependent competence streptococci exhibited a rate of evolution of the SAM domain almost identical to that of the eRF domain (Fig. 5*C*). Because the plot of eRF versus species distances revealed a similar evolution rate for all studied taxonomic groups, we concluded that the SAM domain of ComRS-dependent competence streptococci exhibits a truly reduced evolution rate compared with the SAM domain from other species. This finding is consistent with DprA proteins from this subgroup of streptococci evolving as a unique entity, in contrast to other members of the DprA family. To account for this peculiarity, we propose that these streptococci share DprA proteins that have acquired a unique function involving both domains; hence, they forced their coevolution.

Comparison of ComE and SAM-DprA Phylogenetic Distances. To complete our phylogenetic analysis, we wished to compare the rate of evolution of ComE proteins with that of the SAM domain of DprA. Computation of ComE phylogenetic distances versus DprA eRF phylogenetic distances revealed that ComE proteins evolved as rapidly (slope = 2.56) as their corresponding SAM (Fig. S4*A*). This result is consistent with a coevolution of the two interacting entities and suggests that DprA-dependent shut-off of competence involving direct interaction with ComE is a property presumably shared by the entire subgroup of streptococci with ComE regulated competence.

Notably however, the *Streptococcus mitis* group displayed striking ComE sequence conservation (slope = 0.35) mainly resulting from large differences between SAM and ComE tree topologies and branch lengths for *Streptococcus peroris* and *Streptococcus infantis* (Fig. S4*B*). We interpret the striking difference in the phylogeny of *S. peroris* and *S. infantis* ComE, which contrasts with the homogeneous eRF, SAM, and species phylogenies, as resulting from the horizontal transfer of ComE from a donor species closely related to the *Streptococcus pneumoniae-S. mitis-Streptococcus sanguinis- Streptococcus oralis* subgroup. It is worth mentioning that ComD phylogeny strictly follows that of ComE, suggesting ComD was simultaneously transferred with ComE.

Horizontal transfer would also readily account for the apparent differential behavior in terms of phylogenetic distances of ComE and the SAM of DprA specifically observed for *S. peroris* and *S. infantis* (Fig. S4 *B* and *C*).

SI Materials and Methods

Strains, Culture, and Transformation Conditions. Stock cultures were routinely grown at 37 °C in Todd-Hewitt medium + 0.5% Yeast extract (THY) or C+Y medium to OD_{550} = 0.4; after addition of 15% (vol/vol) glycerol, stocks were kept frozen at −70 °C. For the study of spontaneous or competence-stimulating heptadecapeptide (CSP)-induced competence, cells were incubated in C+Y medium. Unless otherwise indicated, fresh cultures were first grown to OD_{550} = 0.3–0.4, then diluted to OD_{550} = 0.04–0.07 (depending on the experiment) and synthetic CSP1 (100 ng/mL^{-1}, unless otherwise indicated) was added after 10- or 12-min incubation, depending on the experiment.

For chromosomal transformation, DNA was added 10 min after CSP and cells were incubated for 20 min at 30 °C. Transformants were selected by plating on CAT-agar supplemented with 4% (vol/vol) horse blood, followed by selection using a 10-mL overlay containing chloramphenicol (Cm; 4.5 μg/mL^{-1}), erythromycin Ery; (0.05 μg/mL^{-1}), kanamycin (Kan; 250 μg/mL^{-1}), spectinomycin (Spc; 100 μg/mL^{-1}), or streptomycin (Sm; 200 μg/mL^{-1}), after phenotypic expression for 120 min at 37 °C.

Monitoring of *com* Gene Expression. Construction of the transcriptional fusions involved the cloning of a *S. pneumoniae* DNA fragment upstream of the *luc* gene in a nonreplicative plasmid, followed by homology-dependent integration of the recombinant plasmid into the pneumococcal chromosome (1). At each locus, the *luc* gene was thus inserted at a very similar position, next to the 13th or 14th codon of the targeted *com* gene with *luc* translation relying on the same SD; a fully functional copy of the *com* gene was maintained downstream of the integrated plasmid (1). Luciferase activity was directly measured in cultures of pneumococci actively growing in medium containing its substrate, luciferin, as previously described (2). RLU (relative luminescence unit; and when indicated $OD_{492\ nm}$) values were recorded throughout incubation at 37 °C into a 96-well white microplate with clear bottom in a Lucy I luminometer (Anthos) or a Varioskan Flash luminometer (Thermo 399 Electron Corporation). RLU should therefore be compared within but not between figures.

Ectopic Expression of DprA as an Early *com* Gene. To achieve expression of *dprA* as an early *com* gene, $pCEP_E$, an integrative plasmid derived from $pCEP_{cin}$ (3), allowing chromosomal integration of a gene at CEP and its expression under the control of the CSP-inducible, ComE-dependent promoter (P_E) of the *comCDE* operon was constructed. P_E is present on a 172-bp fragment, also containing *comC* RBS [positions 2,035,465–2,035,636 in the R6 genome (4), amplified using the comC-CEP and comCNCO primers (Table S2), and R6 genomic DNA as template]. The resulting PCR fragment was digested with XhoI and NcoI, and inserted into XhoI-NcoI–digested pCEP*cin* yielding plasmid $pCEP_E$. The *dprA* gene was then amplified with primer pair dprA16-dprA18, using R800 chromosomal DNA as template, digested with NcoI/BamHI, and ligated to NcoI/BamHI-digested $pCEP_E$ plasmid. Transformation of strain R1501 with the ligation mixture and selection for Kan^R transformants generated strain R1995, which harbors the CEP_E-*dprA* construct.

Random Mutagenesis and Selection for DprA Interaction-Defective Mutants. The *dprA* coding sequence was first subjected to random mutagenesis by error-prone PCR favoring single mutational events as previously described (5). A library of the mutated DprA coding sequence, fused with the binding-domain functional domains of GAL4, was established in the yeast PJ69-4 (α) haploid strain using the gap-repair procedure and arrayed in the 96-well plate format. An array of 1,500 colonies expressing potential DprA mutant proteins was mated with PJ69-4 (α) strains expressing either ComE, DprA, or RecA proteins in fusion with the complementary GAL4 functional domain. Diploids were monitored for their ability to grow on selective media lacking uracil, leucine, histidine or adenine (−LUH and −LUA). Diploid colonies that failed to grow on selective media were screened, looking specifically for diploids that failed to express the interaction phenotype when DprA was coexpressed with ComE, yet keeping the ability to interact with DprA and RecA. The corresponding haploid clones

harboring the DprA mutant derivatives were pooled and tested again for the loss of interaction with ComE phenotype. The corresponding mutations within *dprA* were identified by sequencing.

Screening for *mariner* Insertions Suppressing *dprA-cup* Synthetic Lethality. In vitro *mariner* mutagenesis of *S. pneumoniae* R800 chromosomal DNA was used to generate a library of ~120,000 minitransposon insertions (conferring resistance to kanamycin, Kan^R) in strain R800, as previously described (6). The Kan^R library was then transferred by transformation into the *trt1* (*comD*^{D299N}) strain, R394, and transformed with *dprA*::*spc21*^{C} chromosomal DNA. From 57 Spc^R transformants isolated, each Kan^R minitransposon was retransformed into the *trt1* strain and the resulting strains were used as recipient for *dprA*::*spc21*^{C}, to establish which of the Kan^R insertions acted as suppressor of synthetic lethality. Seventeen true suppressors were thus validated, of which 15 corresponded to minitransposon insertions in the *comAB* operon (Fig. S3*B*); the remaining two corresponded to complex insertions, one of which resulting in a duplication *dprA*. It is of note that the transformation step required for introduction of the *dprA* knock-out in the *trt1* minitransposon library precluded the recovery of CSP nonresponsive mutants (i.e., mutations abolishing transformation), such as those resulting from insertions in *comD*, *comE*, *comW*, or *comX*.

Construction of a Phylogenetic Tree of Streptococci. To construct a phylogenetic tree of streptococcal species, instead of relying on 16S ribosomal RNA sequences (7), which are poorly discriminant for closely related species, we used Clusters of Orthologous Groups (COG) families. Eighteen complete genome entries were retrieved (see *Building of Sequence Samples*, below) from the European Bioinformatics Institute (www.ebi.ac.uk/genomes/bacteria.html) as well as 18 whole-genome shotgun entries (www.ebi.ac.uk/genomes/wgs.html) for a best-coverage of Streptococcaceae and the *Lactococcus lactis* genome to be used as an outgroup to root the *Streptococcus* tree. We used RPSBLAST program (8) to functionally annotate the whole-protein set of each complete genome with the COG profiles downloaded from the National Center for Biotechnology Institute Conserved Domain Database repository (9). We selected COG families according to the quality of the alignments (E-value < $1e^{-10}$ and an alignment coverage of at least 80% of the COG profile). For each COG family, to ensure that selected sequences were orthologous, we performed reciprocal BLASTP sequence similarity searches to ascertain that they were mutually reciprocal best hits. For the identification of orthologous genes in whole shotgun genomes, the protein sequence of *S. pneumoniae* of each COG family was used as query in TBLASTN (10) sequence-similarity searches against the DNA sequence contigs of each genome. When it occurred, the best-target hit sequence was then used as query in a reciprocal BLASTP searches against the *S. pneumoniae* proteome. If the best hit was the initial query, then we retained the target sequence as a member of the COG family; otherwise, we concluded that the target genome had no record in the COG family. At this step, we retained COG family that cover at least 34 of the 36 genomes.

The alignments for each selected COG family were created using the MUSCLE program (11) with the default parameters. To remove spurious sequences and poorly aligned positions, we used the trimAl program (12) to analyze the quality of the alignments according to gap numbers and residue conservation in the aligned columns. We used the automated parameters recommended by the authors to reconstruct maximum-likelihood trees. At this step, 309 families were selected with minimum identity frequency of 0.6 and minimum frequency of sites without gaps of 0.7. The maximum-likelihood trees were computed with PhyML (13). We used the ProtTest3 program (14) to select the optimal combination of parameters. The most frequent combination was the LG model of sequence evolution with a Γ-correction (four categories of evolutionary rates) and shape parameter and proportion of invariant sites estimated from data. When a species did not have a record for a COG family, the missing sequence was replaced by gaps in the alignment. The aligned sequences of 279 COG families for the 36 species were concatenated together to produce a single alignment of 87,218 aligned positions including 40,635 sites without polymorphism (46.59%). The concatenated tree was computed with PhyML, with the parameters selected by ProtTest3. We statistically evaluated the branch support of the obtained tree with two parametrics (aLRT and SH-like) (15), and one nonparametric (bootstrap with 120 replicates) methods as implemented in PhyML. We also used random removal families resampling (jackknife procedure) as an evaluation of the effect of family selection on the quality of the tree (100 replicates and 100 families selected at each iteration).

Construction of DprA (SAM and eRF), ComE, and ComD Trees, and Computation of Evolutionary Distances. DprA sequences from streptococci, Lactobacillaceae, and *Leuconostoc* species were retrieved as described above. SAM, eRF, ComE, and ComD sequences were aligned with the MUSCLE program (11) and trees were computed with PhyML (13) with the parameters selected by ProtTest3 (14). Evolutionary distances of protein or protein domains between species were computed from sequence alignments with TREE-PUZZLE program (16) using the WAG substitution model, a four γ-distributed rates and the γ-parameter estimated from datasets. For each species pair, the distance computed on the SAM, ComE, or ComD sequences was plotted against the distance computed on the eRF domain of DprA (Fig. 5*C* and Fig. S4*A*). Linear regressions and correlation coefficients were computed with the R package. To discriminate between an acceleration of evolution rate of the SAM domain among streptococci with ComE-regulated competence and a deceleration among streptococci with ComRS-regulated competence, we used the sequences of SAM and eRF domains from two related clades (Lactobacillaceae and *Leuconostoc* species) as control.

Building of Sequence Samples. Whatever the set of analyzed genomes and the protein family (DprA, ComE, or ComD), we used the same procedure. We first retrieved homologous sequences through similarity searches using the corresponding protein sequence from *S. pneumoniae* as query either with the BlastP program of the BLAST suite (8) on all protein sequences deduced from the genome annotation, or with the TBLASTN program against the DNA contigs in case of shotgun genomes. To identify orthologous sequences, each retrieved sequence was then used as query in reciprocal BlastP searches against the complete set of *S. pneumoniae* proteins. If the best hit was the initial *S. pneumoniae* query, we verified that no paralogs either in *S. pneumoniae* genome or in the analyzed genome had a better BlastP similarity score.

1. Martin B, et al. (2013) ComE/ComE~P interplay dictates activation or extinction status of pneumococcal X-state (competence). *Mol Microbiol* 87(2):394–411.
2. Prudhomme M, Claverys JP (2007) *The Molecular Biology of Streptococci*, eds Hakenbeck R, Chhatwal GS (Horizon Scientific, Norfolk, UK), pp 519–524.
3. Martin B, et al. (2010) Expression and maintenance of ComD-ComE, the two-component signal-transduction system that controls competence of *Streptococcus pneumoniae*. *Mol Microbiol* 75(6):1513–1528.
4. Hoskins J, et al. (2001) Genome of the bacterium *Streptococcus pneumoniae* strain R6. *J Bacteriol* 183(19):5709–5717.
5. Noirot-Gros MF, et al. (2006) Functional dissection of YabA, a negative regulator of DNA replication initiation in *Bacillus subtilis*. *Proc Natl Acad Sci USA* 103(7):2368–2373.
6. Martin B, Prudhomme M, Alloing G, Granadel C, Claverys JP (2000) Cross-regulation of competence pheromone production and export in the early control of transformation in *Streptococcus pneumoniae*. *Mol Microbiol* 38(4):867–878.
7. Kawamura Y, Hou XG, Sultana F, Miura H, Ezaki T (1995) Determination of 16S rRNA sequences of *Streptococcus mitis* and *Streptococcus gordonii* and phylogenetic relationships among members of the genus *Streptococcus*. *Int J Syst Bacteriol* 45(2):406–408.

8. Altschul SF, et al. (1997) Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucleic Acids Res* 25(17):3389–3402.
9. Marchler-Bauer A, et al. (2011) CDD: A Conserved Domain Database for the functional annotation of proteins. *Nucleic Acids Res* 39(Database issue):D225–D229.
10. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. *J Mol Biol* 215(3):403–410.
11. Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* 32(5):1792–1797.
12. Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T (2009) trimAl: A tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* 25(15):1972–1973.
13. Guindon S, et al. (2010) New algorithms and methods to estimate maximum-likelihood phylogenies: Assessing the performance of PhyML 3.0. *Syst Biol* 59(3): 307–321.
14. Darriba D, Taboada GL, Doallo R, Posada D (2011) ProtTest 3: Fast selection of best-fit models of protein evolution. *Bioinformatics* 27(8):1164–1165.
15. Anisimova M, Gascuel O (2006) Approximate likelihood-ratio test for branches: A fast, accurate, and powerful alternative. *Syst Biol* 55(4):539–552.
16. Schmidt HA, Strimmer K, Vingron M, von Haeseler A (2002) TREE-PUZZLE: Maximum likelihood phylogenetic analysis using quartets and parallel computing. *Bioinformatics* 18(3):502–504.

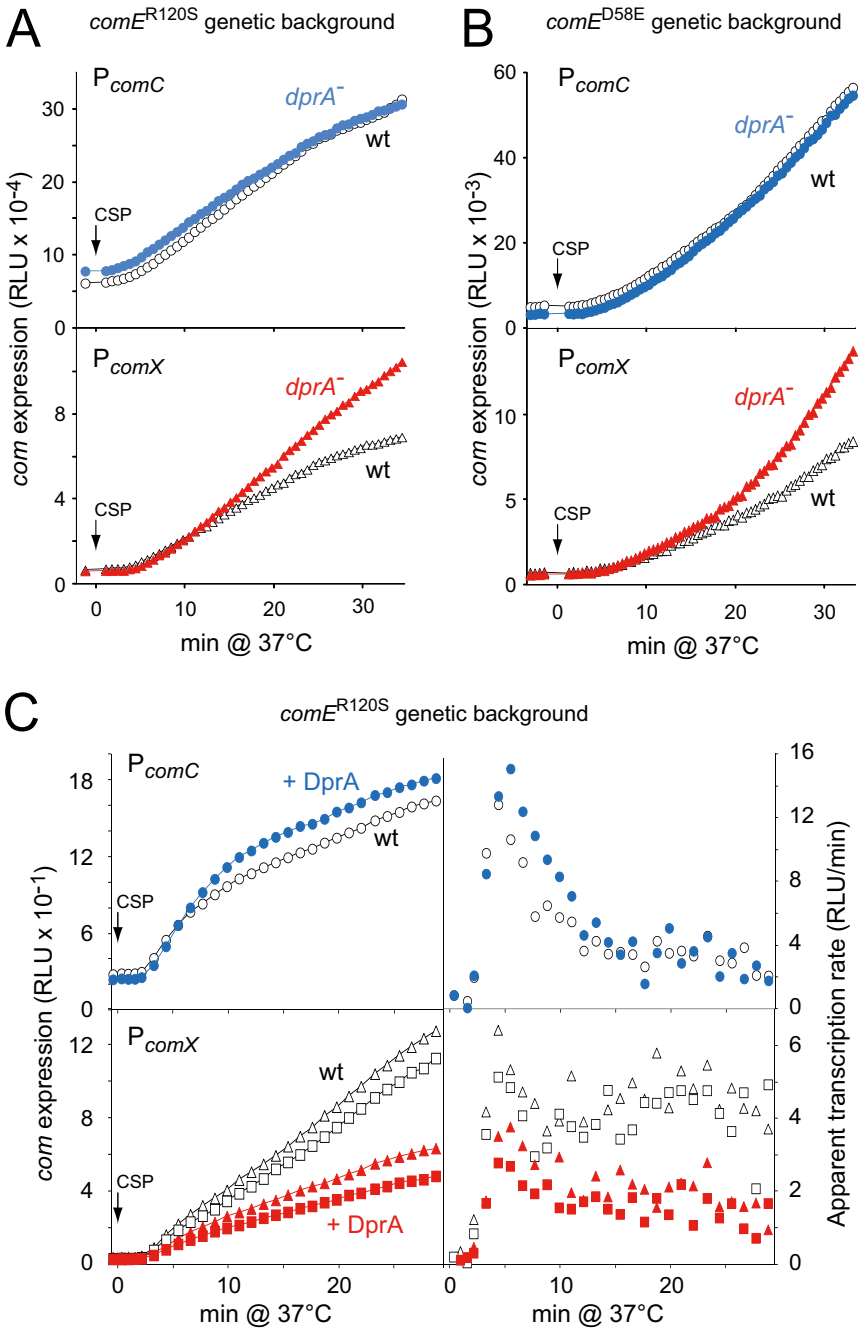

Fig. S1. Inactivation of *dprA* affects mainly *comX* expression in both *comE*^{R120S} and *comE*^{D58E} cells. (*A*) Comparison of P_{comX} and P_{comC} transcription in *comE*^{R120S} cells with and without *dprA* (see Fig. 2*B* legend). *comE*^{R120S} strains used: *comC*::*luc* R2342 (*dprA*$^-$) and R2329 (*Upper*); *comX2*::*luc* R2340 (*dprA*$^-$) and R2314 (*Lower*). (*B*) Comparison of P_{comX} and P_{comC} transcription in *comE*^{D58E} cells with and without *dprA* (see Fig. 2*B* legend). *comE*^{D58E} strains used: *comC*::*luc* R2561 (*dprA*$^-$) and R2446 (*Upper*); *comX1*::*luc* R2559 (*dprA*$^-$) and R2403 (*Lower*). (*C*) Comparison of P_{comX} and P_{comC} transcription in *comE*^{R120S} cells with and without *dprA* expressed as an early *com* gene (see *A*). *comE*^{R120S} strains used: *comC*::*luc* R2341 (CEP_E-*dprA*) and R2329 (*Upper*); *comX*::*luc* R2324-R2325 (CEP_E-*dprA*) and R2313-R2314 (*Lower*).

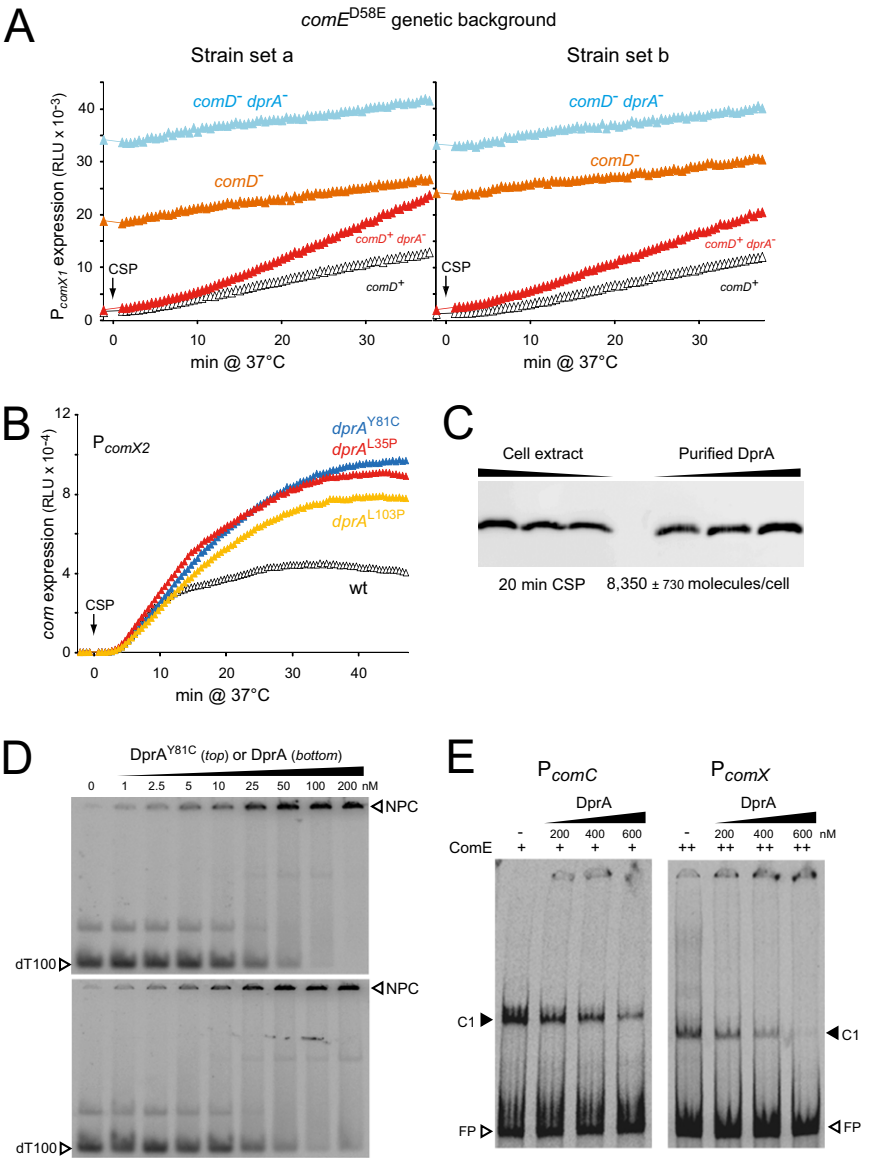

Fig. S2. DprA-ComE interaction is crucial for competence shut-off. (*A*) Evidence that DprA interacts with ComE^{D58E} in competent pneumococcal cells and sequesters it, thus limiting P_{comX} transcription. (*Left* and *Right*) Duplication of the experiment with a different set of strains. Although constitutive high-level expression of *comX1* is observed in *comE*^{D58E} cells lacking ComD (strains R2588 and R2590, *Left* and *Right*, respectively), inactivation of *dprA* results in a further increase in expression (strains R2623 and R2624, *Left* and *Right*, respectively). *comX1* expression levels achieved after CSP addition in *comE*^{D58E} *comD*$^{+}$ (strains R2403 and R2404, *Left* and *Right*, respectively) and *comE*^{D58E} *comD*$^{+}$ *dprA*$^{-}$ (strains R2559 and R2560, *Left* and *Right*, respectively) cells are shown for comparison. CSP addition presumably induces the relief of ComE^{D58E} sequestration by ComD (1). (*B*) DprA mutations abolishing interaction with ComE in yeast two-hybrid (Y2H) affect the shut-off of *comX* expression. Same experimental set-up as in Fig. 3*D*. *comX2*::*luc* strains used: R2200 (*dprA*$^{+}$), R2508 (*dprA*^{L103P}), R2509 (*dprA*^{L35P}), and R2510 (*dprA*^{Y81C}). (*C*) Western-blot analysis of DprA cellular content. Volumes of CSP-induced R1501 extract prepared 20 min after CSP addition, from left to right: 2.5, 1.87, and 1.25 μL. Amounts of purified DprA, from left to right: 2.5, 3.75, and 5 ng; volumes of R2017 (*dprA*-null) extract (prepared as described for R1501 with CSP) from left to right: 1.25, 1.87, and 2.5 μL. Calculations resulted in an estimate of 6,320 ± 2,830 and 8,350 ± 730 (at 20 min) DprA molecules per cell, respectively, 10 min and 20 min after CSP addition. (*D*) EMSA of DprA and DprA^{Y81C} binding to a ^{32}P-dT100 probe. Increasing amounts of purified proteins were incubated with 0.1 nM probe as previously described (2). (*E*) DprA inhibits the binding of ComE to P_{comC} and P_{comX} promoters in vitro. Binding was assayed as described in *Materials and Methods*, except that the indicated concentrations of DprA protein were added to 400 (P_{comC} promoter fragment) or 750 (P_{comX} promoter fragment) nM of ComE. C1, type 1 complex; FP, free probe.

1. Martin B, et al. (2013) ComE/ComE~P interplay dictates activation or extinction status of pneumococcal X-state (competence). *Mol Microbiol* 87(2):394–411.
2. Mortier-Barrière I, et al. (2007) A key presynaptic role in transformation for a widespread bacterial protein: DprA conveys incoming ssDNA to RecA. *Cell* 130(5):824–836.

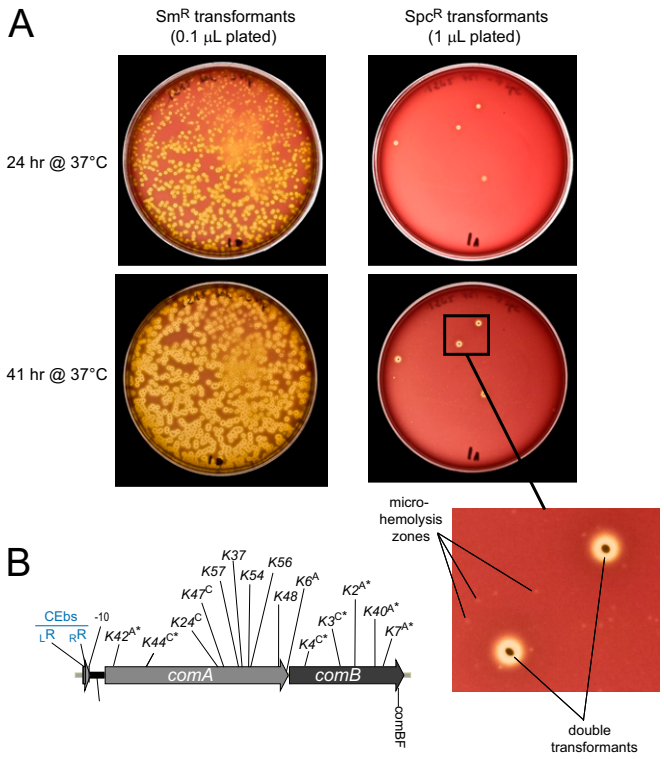

Fig. S3. Synthetic lethality of *dprA* and *comD*^{D299N} *cup* mutations. (*A*) Illustration of the synthetic lethality of a *dprA* knock-out and the *comD*^{D299N} *cup* mutation. Potential *dprA*::*spc* transformants (Spc^{R}) appeared in the *trt1* (*comD*^{D299N}) strain, R394, only after prolonged incubation (41 h) and as microhemolysis zones (see enlargement), which did not grow further and could not be recovered for growth in liquid medium. Spc^{R} transformants (*Right*) with normal colony size (see 24-h plates) appeared with an ~3-log reduced frequency compared with the frequency for the reference marker *rpsL41* conferring resistance to streptomycin (Sm^{R}; *Left*). Genetic analysis revealed that these Spc^{R} colonies corresponded to double transformants having simultaneously inserted the *dprA* knock-out and the wild-type allele at the *cup* site. The 8.5-cm diameter Petri dishes were used for plating. (*B*) Recovery of *mariner* insertions at the *comAB* locus allowing the recovery of *dprA*$^{-}$ transformants in the *comD*^{D299N} recipient. Primer comBF (indicated below the map) and the *mariner* transposon specific primers MP17 or MP128 (Table S2) were used to localize insertions. Superscript C and A indicate cotranscribed and reverse orientation of *kan* (*K*) cassette, respectively, with respect to *comAB*. An asterisk indicates insertions located through DNA sequencing (using primer MP128) at positions +99, +484, +2354, +2762, +2942, +3178, and +3285 for *K42*^{A}, *K44*^{C}, *K4*^{C}, *K3*^{C}, *K2*^{A}, *K40*^{A}, and *K7*^{A}, respectively. Positions are given with respect to the first nucleotide of *comA* taken as +1. CEbs, ComE-binding site consisting of two 9-bp imperfect repeats ($_{L}$R and $_{R}$R) separated by a stretch of 12 nucleotides; the −10 box is located 32 bp downstream of $_{R}$R (1). RR3 indicates the location of a repeated element, BOX, previously identified immediately upstream of *comA* (2, 3).

1. Martin B, et al. (2013) ComE/ComE~P interplay dictates activation or extinction status of pneumococcal X-state (competence). *Mol Microbiol* 87(2):394–411.
2. Martin B, et al. (1992) A highly conserved repeated DNA element located in the chromosome of *Streptococcus pneumoniae*. *Nucleic Acids Res* 20(13):3479–3483.
3. Knutsen E, Johnsborg O, Quentin Y, Claverys JP, Håvarstein LS (2006) BOX elements modulate gene expression in *Streptococcus pneumoniae*: Impact on the fine-tuning of competence development. *J Bacteriol* 188(23):8307–8312.

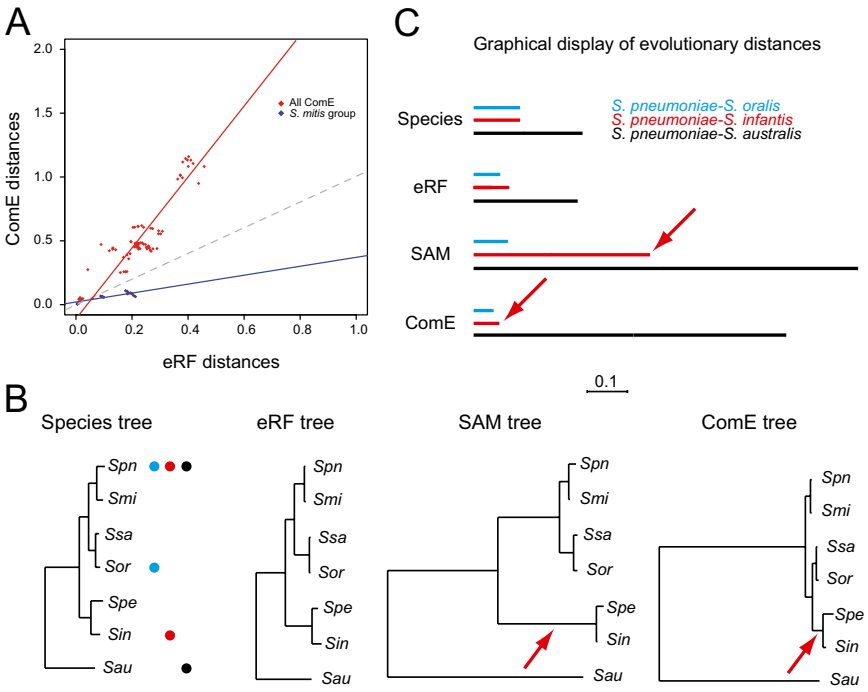

Fig. S4. Comparative evolution of ComE, SAM, and eRF domains among streptococci with ComE-regulated competence. (*A*) ComE from species belonging to the *S. mitis* clade display striking sequence conservation (slope = 0.35; $r = 0.85$) but other ComE evolve as rapidly (slope = 2.56; $r = 0.87$) as the DprA SAM in the same species, with respect to the eRF domain (see Fig. 5*C*). Color code: red, all streptococci with ComE-regulated competence; blue, *S. mitis* clade comprising *S. infantis*, *S. peroris*, *S. oralis*, *S. sanguinis*, *S. mitis*, and *S. pneumoniae* (i.e., species corresponding to digits 9–14 in Fig. 5*A*). *Streptococcus australis* was used as an outgroup. Colored solid lines represent correlation curves for each group of species. Gray dotted line indicates the line of slope 1, expected if ComE and eRF evolved at identical rates. (*B*) Comparison of eRF, SAM, and ComE trees with the phylogenetic tree for the *S. mitis* group. *S. pneumoniae* (*Spn*), *S. mitis* (*Smi*), *S. sanguinis* (*Ssa*), *S. oralis* (*Sor*), *S. peroris* (*Spe*), *S. infantis* (*Sin*), and *S. australis* (*Sau*). These species correspond to digits 14–8 (in descending order) in Fig. 5*A*. Colored and black dots identify pair-wise comparisons for that are graphically displayed in *C*. Red arrows point the striking differences existing between SAM and ComE tree topology and branch lengths for *S. peroris* and *S. infantis*. Distance scale is indicated. (*C*) Graphical display allowing a comparison of species, eRF, SAM, and ComE evolutionary distances between *S. pneumoniae* and *S. oralis* (blue line), *S. pneumoniae* and *S. infantis* (red line), and *S. pneumoniae* and *S. australis* (black line). Red arrows point the striking differences existing between SAM and ComE tree topology and branch lengths for *S. peroris* and *S. infantis*. Same distance scale as in *B*.

Table S1. Correspondence between streptococcal species names, digits, and genome accession numbers

Digit	Streptococcal species* (strain name, when relevant)	Accession number (complete genome)† or whole-genome shotgun entry‡
1	*Streptococcus intermedius*	AJKN
2	*Streptococcus constellatus*	AICQ
3	*Streptococcus anginosus* (strain CCUG_39159)	AICP
4	*Streptococcus anginosus* (strain 1_2_62CV)	ADME
5	*Streptococcus gordonii* (Challis)	NC_009785.1
6	*Streptococcus sanguinis* (strain SK36)	NC_009009.1
7	*Streptococcus cristatus*	AEVC
8	*Streptococcus australis*	AEQR
9	*Streptococcus infantis*	AFNN
10	*Streptococcus peroris*	AEVF
11	*Strepotococcus oralis*	NC_009785.1
12	*Streptococcus sanguinis* (strain ATCC_49296)	AEPO
13	*Streptococcus mitis* (strain B6)	NC_013853.1
14	*Streptococcus pneumoniae* (strain R6)	NC_003098.1
15	*Streptococcus mutans* (strain UA159)	NC_004350.2
16	*Streptococcus macacae*	AEUW
17	*Streptococcus criceti*	AEUV
18	*Streptococcus downei*	AEKN
19	*Streptococcus salivarius*	NC_017595.1
20	*Streptococcus thermophilus*	NC_008532.1
21	*Streptococcus vestibularis*	AEVI
22	*Streptococcus equinus*	AEVB
23	*Streptococcus gallolyticus*	NC_015215.1
24	*Streptococcus pasteurianus*	NC_015600.1
25	*Streptococcus bovis*	AEEL
26	*Streptococcus agalactiae*	NC_004116.1
27	*Streptococcus urinalis*	AEUZ
28	*Streptococcus parauberis*	NC_015558.1
29	*Streptococcus uberis*	NC_012004.1
30	*Streptococcus porcinus*	AEUU
31	*Streptococcus ictaluri*	AEUX
32	*Streptococcus equi*	NC_011134.1
33	*Streptococcus dysgalactiae*	NC_012891.1
34	*Streptococcus pyogenes*	AE004092.1
35	*Lactococcus lactis*	AE005176.1
36	*Streptococcus suis*	CP000407.1

Color code: red, species with ComE-regulated competence; blue, species with ComRS-regulated competence (see Fig. 5*A*); green, identifies *Lactococcus lactis* (used as an outgroup).

**Lactococcus lactis* (#35) excepted.

†Available at www.ncbi.nlm.nih.gov/genome/browse/.

‡Available at www.ebi.ac.uk/genomes/wgs.html.

Table S2. Strains, plasmids, and primers used in this study

Strains, plasmids, or primers	Genotype*/description	Source
S. pneumoniae strains		
R304	R800 derivative, *nov1*, *rif23*, *rpsL41*; Nov^R, Rif^R, Sm^R	(1)
R394	R800, but *trt1* mutation ($ComD^{D299N}$)	(2)
R751	R800 but *dprA*::*spc21*^{C} *rpsL41*; Spc^R, Sm^R	Present study
R800	R6 derivative	(3)
R1245	R394, but *ssbB*::*luc* (pR424), *ssbB*$^+$; Cm^R	Present study
R1501	R800 but Δ*comC*	(4)
R1502	R1501 but *ssbB*::*luc* (pR424), *ssbB*$^+$; Cm^R	(4)
R1800	R1502 but *dprA*::*spc21*^{C}; Spc^R	Present study
R1818	R1501 but *hexAΔ3*::*ermAM*; Ery^R	(5)
R1960	R1501 but *comC*::*luc* (pR428), Δ*comC*; Cm^R	(6)
R1995	R1501 but CEP_E-*dprA*; Kan^R	Present study
R2002	R1960 but *comX1*::*ermAM comX2*::*tet*; Cm^R, Ery^R, Tet^R	(6)
R2003	R1995 but *comX1*::*ermAM comX2*::*tet*; Kan^R, Ery^R, Tet^R	Present study
R2017	R1960 but *dprA*::*spc21*^{C}; Cm^R, Spc^R	Present study
R2018	R1502 but *dprA*::*spc21*^{C}; Cm^R, Spc^R	Present study
R2136	R1818 but *dprA*^{L103P}; Ery^R	Present study
R2141	R1818 but *dprA*^{L35P}; Ery^R	Present study
R2142	R1818 but *dprA*^{Y81C}; Ery^R	Present study
R2149	R1818 but *dprA*^{V31A}; Ery^R	Present study
R2150	R1818 but *dprA*^{Q33R}; Ery^R	Present study
R2178	R2136 but *ssbB*::*luc* (pR424), *ssbB*$^+$; Cm^R, Ery^R	Present study
R2179	R2141 but *ssbB*::*luc* (pR424), *ssbB*$^+$; Cm^R, Ery^R	Present study
R2180	R2142 but *ssbB*::*luc* (pR424), *ssbB*$^+$; Cm^R, Ery^R	Present study
R2181	R2149 but *ssbB*::*luc* (pR424), *ssbB*$^+$; Cm^R, Ery^R	Present study
R2182	R2150 but *ssbB*::*luc* (pR424), *ssbB*$^+$; Cm^R, Ery^R	Present study
R2184	R1818 but *dprA*$^{C234R\text{-}F245L}$; Ery^R	Present study
R2187	R1818 but *dprA*$^{I251V\text{-}H260R}$; Ery^R	Present study
R2189	R1818 but *dprA*$^{D257G\text{-}L269S}$; Ery^R	Present study
R2199	R1501 but *comE*^{R120S}	(6)
R2200	R1501 but *comX2*::*luc* (pR474), *comX*$^+$; Cm^R	(6)
R2209	R2184 but *ssbB*::*luc* (pR424), *ssbB*$^+$; Cm^R, Ery^R	Present study
R2210	R2187 but *ssbB*::*luc* (pR424), *ssbB*$^+$; Cm^R, Ery^R	Present study
R2211	R2189 but *ssbB*::*luc* (pR424), *ssbB*$^+$; Cm^R, Ery^R	Present study
R2218	R1501 but *comX1*::*luc* (pR473), *comX*$^+$; Cm^R	(6)
R2240	R2218 but *dprA*::*spc21*^{C}; Cm^R, Spc^R	Present study
R2241	R2200 but *dprA*::*spc21*^{C}; Cm^R, Spc^R	Present study
R2250	R2218 but *comX2*::*tet*; Cm^R, Tet^R	Present study
R2256	R2250 but *comX1*::*ermAM*; Cm^R, Tet^R; Ery^R	Present study
R2259	R2218 but CEP_E-*dprA*; Cm^R, Kan^R	Present study
R2260	R2200 but CEP_E-*dprA*; Cm^R, Kan^R	Present study
R2265	R2250 but CEP_E-*dprA*, *comX1*::*ermAM*; Cm^R, Tet^R, Kan^R, Ery^R	Present study
R2313	R2199 but *comX1*::*luc* (pR473), *comX*$^+$; Cm^R	(6)
R2314	R2199 but *comX2*::*luc* (pR474), *comX*$^+$; Cm^R	(6)
R2324	R2313 but CEP_E-*dprA*; Cm^R, Kan^R	Present study
R2325	R2314 but CEP_E-*dprA*; Cm^R, Kan^R	Present study
R2329	R2199 but *comC*::*luc* (pR428), Δ*comC*; Cm^R	(6)
R2340	R2314 but *dprA*::*spc21*^{C}; Cm^R, Spc^R	Present study
R2341	R2329 but CEP_E-*dprA*; Cm^R, Kan^R	Present study
R2342	R2329 but *dprA*::*spc21*^{C}; Cm^R, Spc^R	Present study
R2401	R1501 but *comE*^{D58E}	(6)
R2402	R1501 but *comE*^{D58E}	(6)
R2403	R2401 but *comX1*::*luc* (pR473), *comX*$^+$; Cm^R	(6)
R2404	R2402 but *comX1*::*luc* (pR473), *comX*$^+$; Cm^R	(6)
R2446	R2401 but *comC*::*luc* (pR428), Δ*comC*; Cm^R	(6)
R2454	R2136 but *comX1*::*luc* (pR473), *comX*$^+$; Cm^R, Ery^R	Present study
R2455	R2141 but *comX1*::*luc* (pR473), *comX*$^+$; Cm^R, Ery^R	Present study
R2456	R2142 but *comX1*::*luc* (pR473), *comX*$^+$; Cm^R, Ery^R	Present study
R2508	R2136 but *comX2*::*luc* (pR474), *comX*$^+$; Cm^R, Ery^R	Present study
R2509	R2141 but *comX2*::*luc* (pR474), *comX*$^+$; Cm^R, Ery^R	Present study
R2510	R2142 but *comX2*::*luc* (pR474), *comX*$^+$; Cm^R, Ery^R	Present study
R2559	R2403 but *dprA*::*spc21*^{C}; Cm^R, Spc^R	Present study
R2560	R2404 but *dprA*::*spc21*^{C}; Cm^R, Spc^R	Present study

Table S2. Cont.

Strains, plasmids, or primers	Genotype*/description	Source
R2561	R2446 but *dprA*::*spc21*^{C}; Cm^{R}, Spc^{R}	Present study
R2588	R2403 but *comD*::*kan105*^{C}; Cm^{R}, Kan^{R}	(6)
R2590	R2404 but *comD*::*kan105*^{C}; Cm^{R}, Kan^{R}	(6)
R2623	R2588 but *dprA*::*spc21*^{C}; Cm^{R}, Kan^{R}, Spc^{R}	Present study
R2624	R2590 but *dprA*::*spc21*^{C}; Cm^{R}, Kan^{R}, Spc^{R}	Present study
pCEP	pSC101 derivative (i.e., low copy number plasmid) carrying CEP; Kan^{R}, Spc^{R}	(7)
pCEP$_{E}$	pCEP derivative containing instead of the maltose-driven promoter, P$_{M}$, the ComE-dependent promoter, P$_{E}$, of the *comCDE* operon and the RBS of *comC*; Kan^{R}, Spc^{R}	Present study
pKHS	ColE1 derivative; Kan^{R}	(8)
pKHS-*comE*	pKHS derivative carrying the *comE* orf fused to a C-terminal His6 tag; Kan^{R}	(9)
pKHS-*comE*$^{D58E-R120S}$	pKHS derivative carrying the *comE* orf fused to a C-terminal His6 tag; Kan^{R}	(6)
pGKJE3	pACYC184 derivative carrying the chaperone-encoding *groES-groEL-dnaK-dnaJ-grpE* genes under the control of an arabinose-inducible promoter; Cm^{R}	(10)
pR430	pTYB1 derivative, carries a *dprA*-intein fusion, Ap^{R}	(11)
pR502	pR430 derivative, carries a *dprA*^{Y81C}-intein fusion, Ap^{R}	Present study
Primers	Sequence†; gene; position‡	
Oligo-LE	GATAGAGCATTCGCCTTCTAAG; *comC*; −198	(6)
Oligo-RE	GCTACAAACTGTTCCAATTTAAC; *comC*; +12	(6)
comC-CEP	TCGAcTCgaGCTGGGATCAATATAATAGCAAAGCTG; *comC*; −146	Present study
comCNco	CTGTGTTTTcCATggTAAAATCTCCTAAAATG; *comC*; +13	Present study
dprA2	AAATTCCGCAAGAACATCTTGCCCACT; *dprA*; +813	(11)
dprA4	GGAaTTccaTATGAAAATCACAAACTATGAAATC; *dprA*; +24	(11)
dprA16	GAGTTATccATGgAAATCACAAACTATGAAATC; *dprA*; +24	Present study
dprA18	CTAGCTTAGGatccTTTTAAAATTCAAATTCCG; *dprA*; +832	Present study

Superscript C and A indicate, respectively, the cotranscribed and the reverse orientation of an inserted minitransposon antibiotic resistance cassette with respect to the targeted gene.

*Ap, ampicillin; Cm, chloramphenicol; Ery, erythromycin; Kan, kanamycin; Nov, novobiocin; superscript R, resistance; Rif, rifampicin; Spc, spectinomycin; Str, streptomycin.

†Lowercase letters indicate nucleotide different from the *S. pneumoniae* genome sequence, to introduce mutations or convenient restriction sites (underlined).

‡The 3′ oligonucleotide position given with respect to the ATG of the corresponding gene; − and + indicate upstream and downstream, respectively.

1. Mortier-Barrière I, de Saizieu A, Claverys JP, Martin B (1998) Competence-specific induction of *recA* is required for full recombination proficiency during transformation in *Streptococcus pneumoniae*. *Mol Microbiol* 27(1):159–170.
2. Moscoso M, Claverys JP (2004) Release of DNA into the medium by competent *Streptococcus pneumoniae*: Kinetics, mechanism and stability of the liberated DNA. *Mol Microbiol* 54(3): 783–794.
3. Lefèvre JC, Claverys JP, Sicard AM (1979) Donor deoxyribonucleic acid length and marker effect in pneumococcal transformation. *J Bacteriol* 138(1):80–86.
4. Dagkessamanskaia A, et al. (2004) Interconnection of competence, stress and CiaR regulons in *Streptococcus pneumoniae*: Competence triggers stationary phase autolysis of *ciaR* mutant cells. *Mol Microbiol* 51(4):1071–1086.
5. Caymaris S, et al. (2010) The global nutritional regulator CodY is an essential protein in the human pathogen *Streptococcus pneumoniae*. *Mol Microbiol* 78(2):344–360.
6. Martin B, et al. (2013) ComE/ComE~P interplay dictates activation or extinction status of pneumococcal X-state (competence). *Mol Microbiol* 87(2):394–411.
7. Guiral S, et al. (2006) Construction and evaluation of a chromosomal expression platform (CEP) for ectopic, maltose-driven gene expression in *Streptococcus pneumoniae*. *Microbiology* 152(Pt 2):343–349.
8. Eckert K, et al. (2010) The Pih1-Tah1 cochaperone complex inhibits Hsp90 molecular chaperone ATPase activity. *J Biol Chem* 285(41):31304–31312.
9. Boudes M (2011) Etude structurale et fonctionnelle d'acteurs de la transformation génétique naturelle de *Streptococcus pneumoniae*. PhD dissertation (Université Paris-Sud, France).
10. Nishihara K, Kanemori M, Kitagawa M, Yanagi H, Yura T (1998) Chaperone coexpression plasmids: Differential and synergistic roles of DnaK-DnaJ-GrpE and GroEL-GroES in assisting folding of an allergen of Japanese cedar pollen, Cryj2, in *Escherichia coli*. *Appl Environ Microbiol* 64(5):1694–1699.
11. Mortier-Barrière I, et al. (2007) A key presynaptic role in transformation for a widespread bacterial protein: DprA conveys incoming ssDNA to RecA. *Cell* 130(5):824–836.