YraN is a helicase-associated nuclease fostering extended recombination events by natural transformation

Léo Hardy[1], Julie Plantade[1], Violette Morales[2,3], Fanny Mazzamuro[4,5], Eduardo P. C. Rocha[4], Patrice Polard[2,3] and Xavier Charpentier[1,*]

[1] CIRI, Centre International de Recherche en Infectiologie, Inserm, U1111, Université Claude Bernard Lyon 1, CNRS, UMR5308, École Normale Supérieure de Lyon, Univ Lyon, 69100, Villeurbanne, France. Team Horigene.
[2] Laboratoire de Microbiologie et Génétique Moléculaires (LMGM), Centre de Biologie Integrative (CBI), Centre National de la Recherche Scientifique (CNRS), 31062 Toulouse, France.
[3] Université de Toulouse, Université Paul Sabatier, 31062 Toulouse, France.
[4] Institut Pasteur, Université Paris Cité, CNRS UMR3525, Microbial Evolutionary Genomics, Paris, 75015, France.
[5] Collège Doctoral – Sorbonne Université, F-75005 Paris, France – France

* Correspondence to :
xavier.charpentier@univ-lyon1.fr

Running title: YraN, a nuclease involved in natural transformation

Abstract

Natural transformation is a widespread molecular pathway of horizontal gene transfer involving the uptake and recombination of exogenous DNA. Exogenous DNA follows a pathway involving genes sequentially required for its capture, internalization, protection, and recombination with the chromosome. Most of these genes were identified through the isolation of transformation-defective mutants and/or based on their expression preceding natural transformation. Yet, genes required for key steps of the pathway remain elusive. We sought to identify any missing component by comparing Tn-seq data obtained in two distantly-related transformable diderm species, the human pathogen *Legionella pneumophila* and the cyanobacterium *Synechococcus elongatus*. We identified *yraN*, a widespread and highly conserved gene of unknown function required for natural transformation. We provide evidence that YraN is a nuclease associated with the ComM helicase, which cooperate to process the D-loop formed by the invasion of the transforming DNA in the chromosomal DNA strands. We propose a model in which cleavage of the displaced strand by YraN can promote the recombination of transforming DNA, leading to extended recombination events. The identification of this YraN/ComM nuclease/helicase system supports the hypothesis that bacteria possess a conserved pathway for the transport and recombination of exogenous DNA.

Significance

Many bacteria use a dedicated pathway to internalize and integrate extracellular DNA into their chromosome. This allows them to naturally acquire genes, or gene variants, that can confer them new traits, hence the term of natural transformation. Although reported nearly a century ago by Frederick Griffith, several aspects of the mechanism of natural transformation remain elusive. Specifically, it is not clear whether or not all the conserved molecular players of the pathway have been identified. We queried two distinct bacterial species for any gene that would be required for natural transformation. We confirmed all core players of the pathway, but also uncovered a highly conserved one, YraN. We provide evidence that YraN is an endonuclease, which in association with the ComM helicase, promotes the efficient integration of the extracellular DNA in the bacterial chromosome.

Introduction

One of the remarkable features of bacteria is their ability to naturally undergo genetic transformation. This property is linked to the unique ability of bacteria to import exogenous DNA and integrate it into their chromosome by homologous recombination (1). Despite major differences in their cell wall architecture, the process is generally conserved between monoderms (bacteria with a single cell membrane, like most Gram-positive) and diderms (bacteria with an outer membrane like most Gram-negative) (2). The process is initiated by the capture of DNA by an extracellular Type IV filament (3, 4). While in diderms, the Type IV filament serves other functions such as motility and adhesion (5), the competence-associated Type IV filament of monoderms is, to the best of our knowledge, dedicated exclusively to the transformation pathway. In diderms, retraction of the filament conveys DNA through the outer membrane by passing through the PilQ secretin (4). Retraction of the competence pilus has also been observed in the monoderms *Streptococcus pneumoniae* and *Bacillus subtilis*, and is essential for natural transformation (6, 7). In the periplasmic space, the ComEA DNA receptor would function as a ratchet for double-strand DNA (dsDNA) (8, 9). The DNA entry step is less well understood, the dsDNA is converted into single-stranded DNA (ssDNA) while transported across the cytoplasmic membrane, presumably through a channel formed by the ComEC protein (10). In diderms, the ComFC protein is also required for this transport stage, yet the function of this protein remains elusive (11–13) Upon entry in the cytoplasm, the ssDNA is covered by the ssDNA-binding protein DprA to form a nucleofilament which both protects the incoming DNA from the action of nuclease but also promotes RecA-dependent recombination with the chromosome (14, 15). RecA-dependent strand invasion displaces the homologous chromosomal strand, forming a displacement loop (D-loop) onto which the ComM helicase is recruited (16). The helicase activity of ComM is proposed to favor integration of transforming DNA by promoting extension of the D-loop formed by the invasion of the transforming ssDNA in the chromosomal DNA (17). In many bacteria, the expression of the genes involved in this DNA uptake and recombination pathway is concerted, occurring in a physiological state known as competence, and controlled by divergent regulatory pathways (1).

All proteins specific to the pathway have been identified by mutants defective for transformation or by their observed co-regulation in the state of competence-induced (12, 18–21). Yet, as several key steps of the process remain unresolved (ssDNA conversion, energy, mechanism of D-loop resolution) it remains possible that some players could not be identified through these approaches, either because the encoding genes are small (and thus have evaded mutagenesis) or are not part of the competence regulon. Here, we sought to identify any other highly conserved genes involved in

this pathway. We did so through the reanalysis of comprehensive genetic screens for competence development, and which were based on the ability of bacteria to capture exogenous DNA by natural transformation.

Results

YraN, a protein of unknown function, is required for natural transformation

In order to identify conserved genes involved in natural transformation, we compared the results of transposon-insertion sequencing (Tn-seq) screens conducted in two diderm bacteria, the human pathogen *Legionella pneumophila* (22) and the cyanobacterium *Synechococcus elongatus* (23). Both screens were initially designed to identify all genes required for competence (gene regulation) and natural transformation (22, 23). Only 13 genes were found to be required for natural transformation in both species (Table 1). As expected, no genes correspond to regulatory elements, as the regulation of competence largely differs between the two organisms. Consistent with the role of Type IV pilus in natural transformation, 7 genes encode the core components of Type IV pilus, including the components of the basal body (PilM, PilN, PilO, PilP and PilC), the PilB ATPase and the PilQ secretin. 5 other genes encode the highly conserved known players of the DNA transport and processing pathway (ComEC, ComEA, ComFC, DprA and ComM). The last gene, annotated as *yraN*, encodes a protein of unknown function, belonging to the uncharacterized protein family UPF0102. Tn-seq analyses are based on the $\log_2$ fold change in the number of insertions found in a given gene between the transformed and non-transformed populations. This metrics can identify genes required for transformation, but may not constitute accurately report on the amplitude of the transformation defects of the mutants. For instance, mutants of *comEC* and *comM* in *L. pneumophila* show similar $\log_2$ fold change. Yet, the loss of the putative DNA import channel ComEC completely abrogates transformability in all tested species (2), while the loss of the ComM helicase was found to reduce the transformation frequency of single nucleotide polymorphism (SNP) by 7 to 15 fold in *Acinetobacter baylyi* and *Vibrio cholerae*, respectively (17). We thus sought to test the contribution of YraN in natural transformation in two distinct diderms, *L. pneumophila* (*Lp*) and *Acinetobacter baylyi* (*Ab*). In both species, inactivation of *yraN* did not abolish transformation, but lead to a strong reduction (40-fold in *Lp*, 120-fold in *Ab*) in the transformation efficiency using DNA with homology to the chromosome and carrying a selectable SNP (Figure 1A).

Table 1. Genes required for competence and natural transformation in both *L. pneumophila* and *S. elongatus*. Data were retrieved from previous studies (22, 23). CDS were consistently annotated using eggNOG-mapper (24) and orthologous genes in *L. pneumophila* and *S. elongatus* were matched based on gene name and/or COG when available. Values correspond to the log_2 fold change (log_2FC) in abundance of insertions in the corresponding genes between the population obtained following transformation and the control, (non-transformed, population). Data were sorted and filtered for $log_2FC < -1$ and with P-value < 0.05.

Gene name	COG	Annotation	Ratio transformed vs control populations (log2 fold change)	
			L. pneumophila	*S. elongatus*
comM	COG0606	ComM DNA helicase	-3.3	-1.4
comEC	COG0658	DNA transport transmembrane protein ComEC	-2.9	-2.4
comEA	COG1555	Competence protein ComEA	-3.5	-2.5
pilM	COG4972	Tfp pilus assembly protein PilM	-3.0	-2.9
pilN	COG3166	Tfp pilus assembly protein PilN	-3.6	-2.8
pilO	COG3167	Tfp pilus assembly protein PilO	-2.7	-2.6
pilP	COG3168	Tfp pilus assembly protein PilP	-2.6	-2.4
pilQ	COG4796	Tfp pilus assembly protein PilQ	-3.6	-2.7
pilB	COG2804	Tfp pilus assembly protein PilB	-3.7	-3.1
pilC	COG1459	Tfp pilus assembly protein PilC	-3.3	-3.1
comFC	COG1040	Competence protein ComFC	-3.0	-2.0
dprA	COG0758	DNA protection and recombination DprA	-2.7	-2.3
yraN	COG0792	UPF0102 protein, unknown function	-2.8	-1.0

YraN belongs to the PD-(D/E)XK superfamily of phosphodiesterase, with ssDNA endonuclease activity

YraN is a small protein displaying the domain UPF0102. This domain adopts a common fold consisting of four β-sheets sandwiched between two α-helices which form the active site of the large superfamily of PD-(D/E)XK phosphodiesterases (25, 26). Enzymes belonging to this family are very diverse, and while retaining the same structural scaffold for the phosphodiesterase activity, show little sequence conservation outside of the residues involved in the active site. They are found in prokaryotes, eukaryotes and archaea, performing important functions related to nucleic acid maintenance (i.e., restriction enzymes, tRNA splicing endonucleases, Holliday junction resolvase, transposase) (25, 26). Many display variations in the PD-(D/E)XK structural domain, and/or carry additional domains (UPF0102). In contrast, the structure of YraN, predicted by alphafold2 (27), shows that it solely consists of a minimal PD-(D/E)XK structural domain with the canonical signature of the phosphodiesterase active site (Figure 1B). On the *L. pneumophila* YraN, the active site is thus composed of D_{37}, E_{50} and R_{52} (Figure 1B). In order to test its enzymatic activity, the *L. pneumophila* YraN protein and the E50A mutant $YraN^{E50A}$ were purified to homogeneity (Figure 1C). We found that YraN displays a nicking activity on double-stranded DNA, relaxing supercoiled circular DNA. As anticipated for this family of phosphodiesterase, this single-strand endonuclease activity requires the residue E50 (Figure 1D). Importantly, YraN also acts on single-stranded DNA, with highly relaxed site specificity (Figure 1E). In all, we found that YraN is an endonuclease

capable of cleaving single-stranded DNA. We thus tested whether the enzymatic activity of YraN is required for its function in natural transformation. The wild-type YraN, expressed ectopically from a non-repressible promoter on the *L. pneumophila* native large plasmid pLPP, could restore the loss of transformability caused by the deletion of the chromosomal *yraN* gene (Figure 1A). In contrast, mutants of the active site YraN^{E50A} and YraN^{E37A} could not restore transformability (Figure 1A). Identical results were obtained with a FLAG-tagged version of YraN in *A. baylyi*, which could restore the transformation defect of a *yraN* mutant when ectopically expressed from an IPTG-inducible promoter. While expressed at similar lever, the FLAG-tagged mutant YraN^{E60A} (equivalent to E50A in YraN$_{Lp}$) could not restore the transformation defect (Figure 1A). This shows that the endonuclease activity of YraN is central to its contribution to the pathway of natural transformation.

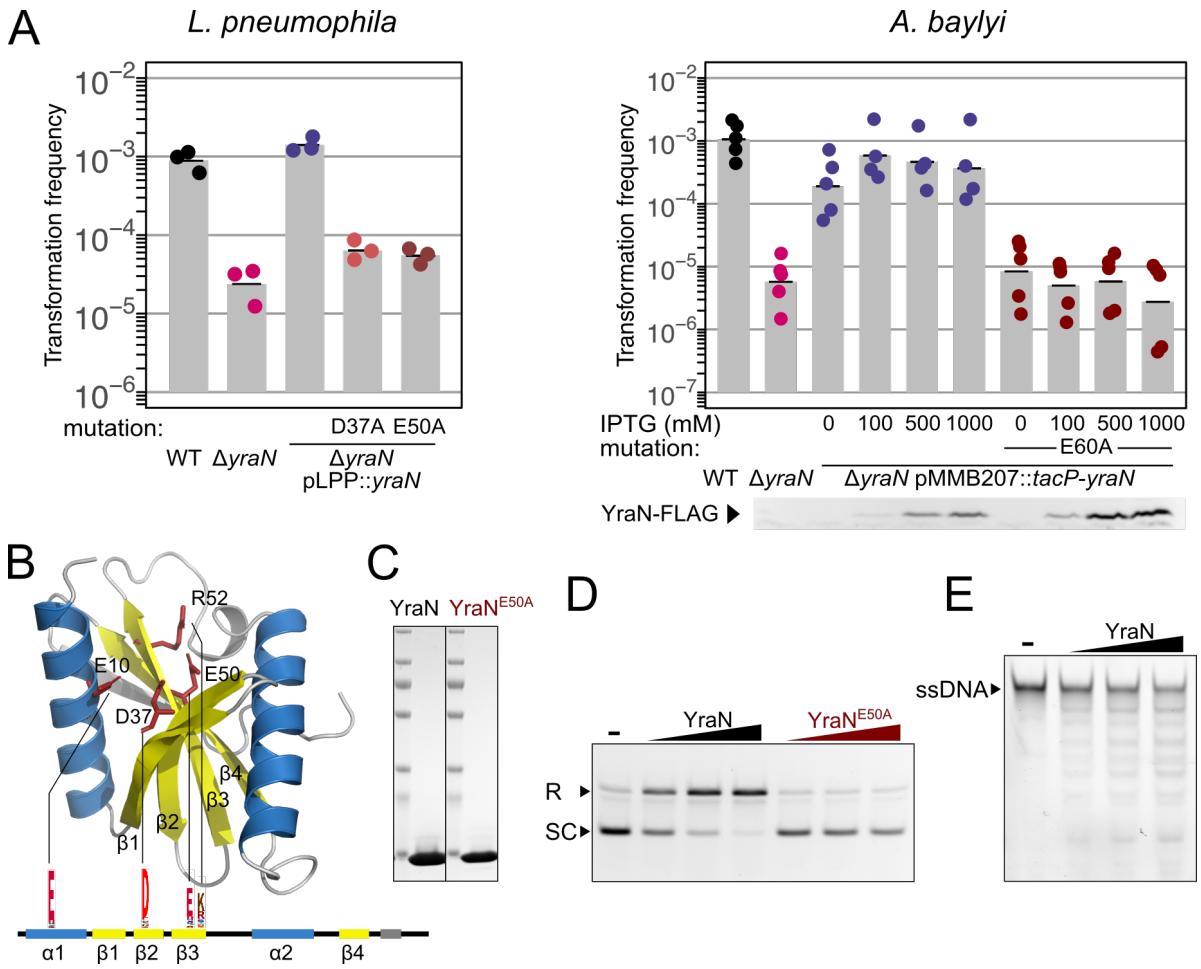

Figure 1. YraN, a protein of unknown function, is required for natural transformation.
A. Efficiency of natural transformation of a SNP conferring resistance to streptomycin (*L. pneumophila, Lp*) or rifampicin (*A. baylyi*, *Ab*). Parent strain (WT), mutants of *yraN*, and complementation clones, were exposed to transforming DNA consisting of a 4 kb-long PCR product encompassing a *rpsL* allele conferring resistance to streptomycin for *Lp* or an 11 kb-long PCR product encompassing an *rpoB* allele conferring resistance to rifampicin for *Ab*. The complementing *yraN* allele in *Lp* is placed in an endogenous plasmid pLPP and expressed from a constitutive promoter. The complementing *yraN-flag* allele in *Ab* is placed in a RSF1010-derived plasmid under control of an IPTG-repressible promoter. D37A and E50A correspond to mutations introduced in the catalytic site of YraN of *Lp*. These residues are found respectively in the beta sheet 2 and 3 (see panel B). E60A in YraN of *Ab* is same mutation as E50A in YraN of *Lp*. Transformation frequencies represent the ratio of colony-forming units on selective (containing

streptomycin or rifampicin) over the total CFU count on non-selective plates (CYE for *Lp*, LB for *Ab*). Parent strain of *Lp* is the constitutively competent *rocC* mutant of the Paris strain. Parent strain of *Ab* is strain BD413. Data represent the geometric average of three (*Lp*) and five (*Ab*) independent experiments.
B. Predicted structure of the *L. pneumophila* YraN using alphaFold2. YraN displays the typical fold of PD-(D/E)XK phosphodiesterase consisting of four beta strands (yellow) sandwiched by who alpha helices (blue). The linear representation of YraN shows the highly conserved residues (red) indicated as weblogo of the UPF0102 domain. The same residues are highlighted in the 3D structure.
C. The *L. pneumophila* YraN and E50A mutant protein were expressed in *E. coli* and purified to homogeneity.
D. Agarose gel electrophoresis of plasmid DNA incubated with increasing amount of YraN and the E50A mutant. Incubation with YraN, but not with the mutant E50A, produces a shift of the supercoiled (SC) fraction to the relaxed (R) form, indicative of ssDNA phosphodiesterase activity.
E. Acrylamide gel electrophoresis of ssDNA incubated with increasing amount of YraN. The ladder below the main ssDNA band reveals endonuclease activity.

YraN is widespread and part of the transformation recombination pathway

In the process of natural transformation, initial cleavage of DNA is thought to occur at the bacterial cell surface, which is then converted into ssDNA as it crosses the cytoplasmic membrane. We considered a possible involvement of YraN at this step. However, the single-strand nicking activity of YraN is neither optimal to cause DNA double-strand cleavage nor to convert dsDNA into ssDNA. This suggested that YraN does not act at the step of DNA uptake/import. In addition, inactivation of *yraN* had no impact on the uptake and establishment of a replicative plasmid (Figure S1A). Thus, *yraN* is not involved in DNA uptake/import but might rather be required for downstream events, possibly to promote chromosomal integration of DNA. To test whether YraN is broadly involved in chromosomal integration of DNA, we tested its involvement in the conjugative transfer of the p45 integrative conjugative element (28) in *L. pneumophila*. We found this site-specific recombination to be insensitive to inactivation of *yraN* (Figure S1B). We then also tested whether YraN might be involved in homologous recombination by analyzing the repair of DNA damage caused by UV irradiation. In contrast to the *recA* mutant, *yraN* mutants of *L. pneumophia* and *A. baylyi* are not defective in the repair of DNA damage caused by UV irradiation (Figure S2).

The previous results led us to propose that YraN has a role in the specific recombination of transforming DNA. Supporting this hypothesis, initial submission of the *L. pneumophila* YraN protein-coding gene to the STRING database returned a network of co-occurrence with genes encoding ComEC, DprA, and, with the highest confidence score, the ComM helicase (Figure S3). We thus investigated further the phylogenetic distribution of YraN, ComM but also DprA and RecA (Figure 2A). RecA is usually regarded as ubiquitous across the bacterial phylogenetic tree and was indeed found in 97.5% of the 4185 queried genomes. DprA is also widespread (92.6%) lacking only in some obligate mutualistic endosymbionts among Bacteroidota, mostly *Blattabacterium*, and *Buchnera* which had previously been described as lacking *recA* (29). The genes *comM* and *yraN* have similar distribution patterns (Figure 2A). They co-occurred in 75% and were both missing in

15% of the genomes. The only large phylum where they were often absent was the Firmicutes, with only 38% of them having both *yraN* or *comM,* mainly Firmicutes from the Clostridia and the Bacilli classes. The presence of *comM* was positively associated with the presence of *yraN* (phyloglm test, effect=4.16, $p=8.25x10^{-4}$).

We then analysed the co-localization of genes in the chromosome because it is often a sign of functional interaction. In 34% of the bacterial genomes having both *comM* and *yraN*, the genes were co-localized in the genome. The proportion of co-localizing *comM* and *yraN* in *comM*+/*yraN*+ genomes varied with the phylum. Despite fewer genomes carrying *yraN* and *comM* in Actinobacteria and Firmicutes, a larger proportion of them, 98% and 66% respectively, had the genes co-localized in the chromosome. The value was substantially lower in Proteobacteria (16% of *yraN*+/*comM*+ genomes). Interestingly, in Corynebacteria the YraN and ComM domains were most frequently found (64/70) forming a YraN-ComM fusion protein.

The bioinformatic analyses indicate a functional association between YraN and the ComM helicase. Indeed, in the two tested species, inactivation of *comM* leads to a transformation defect of SNP similar to that of the *yraN* mutants (Figure 2B). Moreover, these transformation defects are not cumulative, as a *yraN comM* mutant does not show stronger transformation defects than the single mutants (Figure 2B). This clearly suggests that YraN and ComM cooperate in the transformation pathway. The functional association of ComM and YraN is further supported by large-scale co-evolution analyses which predicted the direct interaction of these proteins (30, 31) as well as the interaction between ComM and DprA (30). We thus tested these direct interactions using the bacterial two-hybrid assay based on the functional reconstitution of the *Bordetella pertusis* adenylate cyclase from its T18 and T25 domains fused to the bait and prey proteins (32). We found that, as previously reported, both DprA and RecA could self-interact and interact with one another (Figure 2C). While we detected no interaction between RecA and ComM, we did observe the predicted interaction of ComM and DprA, yet only when ComM is fused to T18 (Figure 2C). Fusing the larger T25 part to the C-terminus of ComM may make it non-functional, as previously reported with fusion to sfGFP (17). We found that YraN could also interact with the ComM fusion which interacts with DprA but also with both T18 and T25 fusions of DprA (Figure 2C). Taken together with phenotypic, phylogenetic, and co-evolution analyses, this experimental interaction assay indicates that YraN is a component of the pathway dedicated to the recombination of single-stranded transforming DNA.

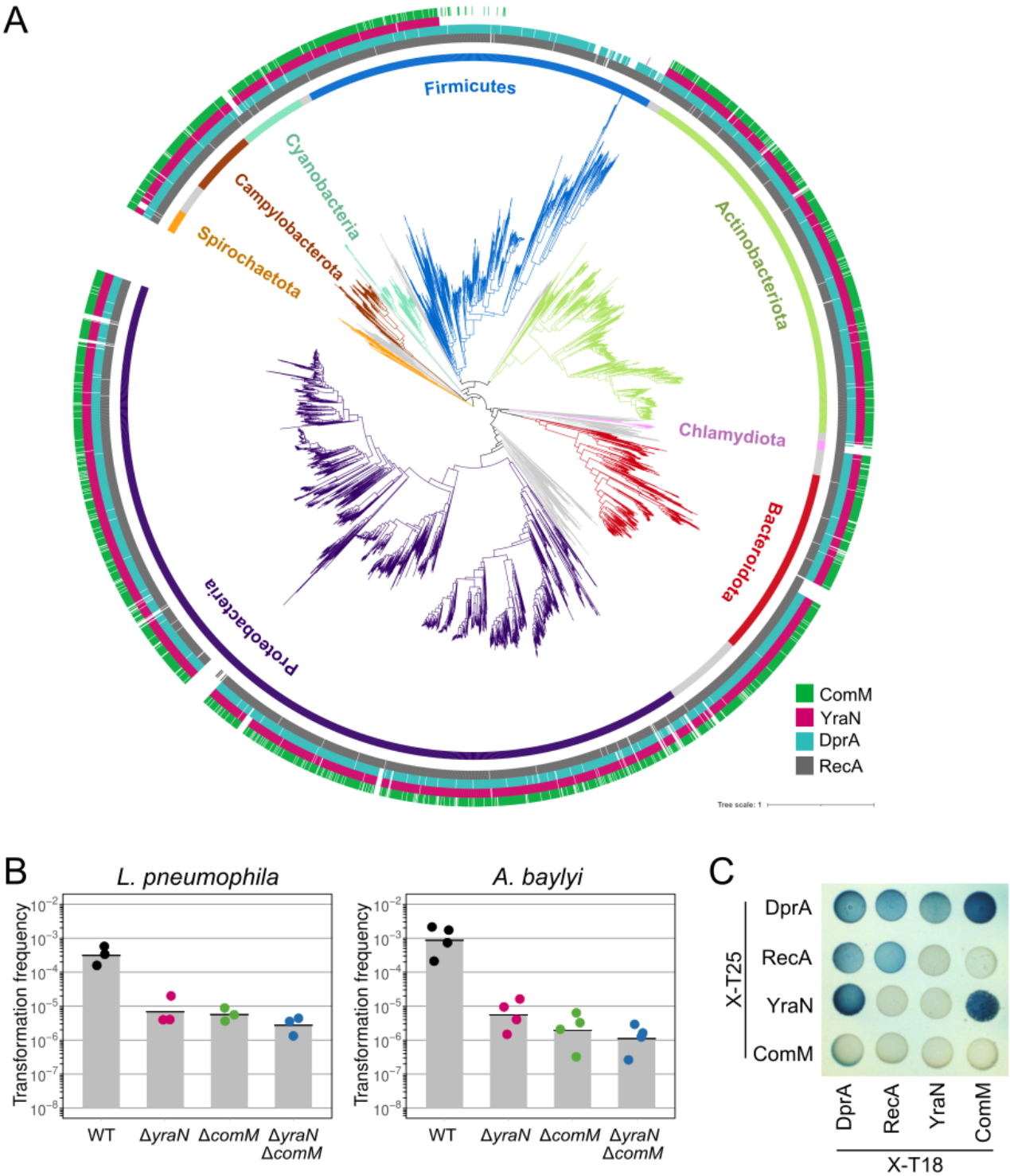

Figure 2. YraN is widespread and part of the transformation recombination pathway.
A. Distribution of selected transformation-associated proteins in the bacterial phylogenetic tree. The tree consists of 4185 strains, out of 21,105 bacterial genomes searched for the presence of YraN, ComM, DprA, and RecA. The taxonomy was provided by Genome Taxonomy Database (GTDB).
B. Efficiency of natural transformation of a SNP conferring resistance to streptomycin (*L. pneumophila, Lp*) or rifampicin (*A. baylyi*, *Ab*). Parent strain (WT), single mutants of *yraN, comM* and double mutants *yraN* and *comM* were exposed to transforming DNA consisting of a PCR product of the *rpsL* allele for *Lp* or *rpoB* allele for *Ab*. Transformation frequencies represent the ratio of colony-forming units on selective (containing streptomycin or rifampicin) over the total CFU count on non-selective plates (CYE for *Lp*, LB for *Ab*). Transformation frequencies are from three to four independent experiments (on different days).

C. Bacterial two-hybrid assay to detect interaction between the *L. pneumophila* YraN, DprA, RecA, and ComM proteins. Each of these proteins was expressed in *E. coli* as a fusion with their C-terminus fused to the T18 of T25 domains (X-T18 or X-T25) of the adenylate cyclase of *B. pertusis*. Protein interactions reconstitute of the adenylate cyclase, to produce cAMP, a cofactor required for the expression of the *lacZ* gene. Expression of LacZ is detected using the chromogenic beta-galactosidase substrate X-Gal. Strains expressing combinations of the fusions were spotted on LB plates containing X-Gal and IPTG. Blue coloration indicates interaction between the tested proteins.

YraN promotes D-loop extension in association with the ComM helicase

The functional association of YraN and ComM suggests that YraN participates in the efficient integration of transforming DNA promoted by ComM. ComM was previously proposed to function as a DNA helicase, recruited at the D-loop formed by the invasion of the single-stranded transforming DNA in the chromosome (16, 17). The helicase activity of ComM is thought to allow extension of the D-loop, which was experimentally observed by the comigration of linked genetic markers (17). To test the hypothesis that YraN might assist ComM in D-loop extension, we determined the efficiency of integration of genetic markers separated by increasing distances. An *A. nosocomialis* strain carrying an integrated tetracycline resistance gene followed by the *lacZ* gene was exposed to transforming DNA encompassing the tetracycline resistance, but replaced by a kanamycin resistance gene, and with a 7-bp deletion in the *lacZ* gene at increasing distance from the kanamycin resistance marker (Figure 3A). Selecting transformation events using the kanamycin marker, we then determined the comigration efficiency of the unselected 7-bp deletion (kan^R $lacZ^-$ CFU vs all kan^R CFU) as a function of the distance separating these markers. In the wild-type background, the comigration efficiency decreases with the distance separating the markers, from nearly 1 at 0.25 kb distance down to 0.7 at 3 kb (Figure 3B). While comigration at the 0.25 kb distance is unchanged in a ComM-deficient strain, it rapidly decreases with increasing distance, down to 0.4 at 3 kb (Figure 3B). This confirms previous results obtained in *V. cholerae* and *A. baylyi* (17). The *yraN* mutant is also severely affected for comigration at 3 kb, similarly to the *comM* mutant, suggesting that *yraN* is required for the ComM-dependent D-loop extension. Yet, at 1.6 kb, the *yraN* mutant is less deficient than the *comM* mutant, suggesting that YraN facilitates ComM-dependent D-loop extension over long distances. Consistent with results from transformation assays, the deficiencies in comigration of the *comM* and *yraN* mutant do not seem additive.

The ComM/YraN helicase/nuclease system promotes integration of large, polymorphic DNA sequence

To further characterize the role of ComM and YraN in integration of transforming DNA, we sought to analyze the recombination events resulting from natural transformation. In order to obtain a high-resolution map of recombination events, we exploited the fact that a sequence with mismatches to

the native homolog (homeologous sequence) is efficiently recombined by the natural transformation pathway. An *A. baylyi* strain was provided with a linear 11.2 kb-long DNA fragment encompassing the *rpoB* gene and obtained from a rifampicin-resistant isolate of *A. nosocomialis*. Gene synteny is maintained in this region which contains essential genes involved in transcription and translation. However, the homeologous sequence differs by 1,081 SNPs which can be used to characterize the recombination events in the *A. baylyi* transformants (Figure 3C and 3D). The transformation frequency using this homeologous DNA (on average, 1 SNP every 10 bp) remained high, nearly identical to that of perfectly homologous DNA. Inactivating *yraN*, *comM* or both, resulted in a transformation defect of homeologous DNA similar to between identical DNA sequences (Figure S4). The strong reduction in transformation frequency in the absence of YraN or ComM indicates that the majority of events that could be produced by a wild-type strain fail to produce a selectable transformant in a *yraN* or *comM* mutant. Hence, the recombination events that could be observed in these mutants represent a fraction of those produced in a wild-type strain. This could represent an inherent bias in comparing the characteristics of the recombination events, underestimating possible differences between wild-type and mutants strains.

We sequenced the targeted region and mapped the acquired SNPs defining the recombination events in 63 to 71 transformants of each genotype (WT, *yraN*, *comM,* and *yraN comM*) (Figure S5). While all transformants have acquired the SNP conferring resistance to rifampicin, the probability of acquiring unselected SNPs decreases with the distance separating them from the selected SNP (Figure 3D). This probability decreases more rapidly for the *yraN*, *comM,* and *yraN comM* mutants than for the wild-type (Figure 3), corroborating the results obtained in the comigration genetic assay (Figure 3A). The characterize the recombination events, we defined them as the sequence separated by the two outermost acquired SNPs. Transformants of the wild-type strain present recombination events with a mean length of 6.7 kb (Figure 3E). Mutants of *yraN* and *comM* show events with a mean length significantly shorter than for the wild-type (one-way ANOVA, $p<0.01$) and of 5.5 kb and 4.3 kb, respectively. Interestingly, and again corroborating the comigration assay, the *comM* and *yraN comM* mutants show identical (p-value>0.9) and stronger defects than the *yraN* mutant (p-value<0.01), suggesting that YraN acts as a facilitator of ComM activity (Figure 3E). Recombination events often presented a mosaic structure, consisting of regions of successively acquired SNPs interrupted by one or more non-acquired SNPs. In a WT strain, a recombination event consists of an average of 3.9 regions (Figure 3F). All mutants showed a higher average number of recombined regions. The *yraN* mutant showed nearly twice as many regions (7.4) (Figure 3E), although this does not seem statistically different from the wild-type (Wilcoxon-Mann-Whitney on log-transformed values, $p=0.1$). Both the *comM* and *comM yraN* display even higher (8.8 and 10.5, respectively), and statistically significant ($p<0.05$), average number of regions

(Figure 3F). Recombined segments, defined as regions of more than successive acquired 2 SNPs, are also significantly shorter (Kruskal-Wallis, Dunn test, p<0.01). Indeed, as for the length of events, the *comM* and *yraN comM* mutants show identical (2.2 and 2.3 kb, respectively, p-value>0.9) and shorter mean segment length than the *yraN* mutant (3.2 kb, p-value<0.01) (Figure 3F). In all, the analysis of recombination events indicates that they differ between the wild-type and mutants in two ways: they are generally shorter and are more fragmented. Although qualitatively affected similarly, recombination events are quantitatively less impacted in a *yraN* mutant than in the *comM* mutant, indicating that ComM could function in the absence of YraN, albeit less efficiently. In contrast, the *comM* and *yraN comM* mutants are indistinguishable, indicating that YraN cannot contribute to DNA integration in the absence of ComM.

Altogether, the genetic comigration assay and the high-resolution analysis of recombination reveal that YraN acts to promote ComM-dependent generation of larger and contiguous recombination events.

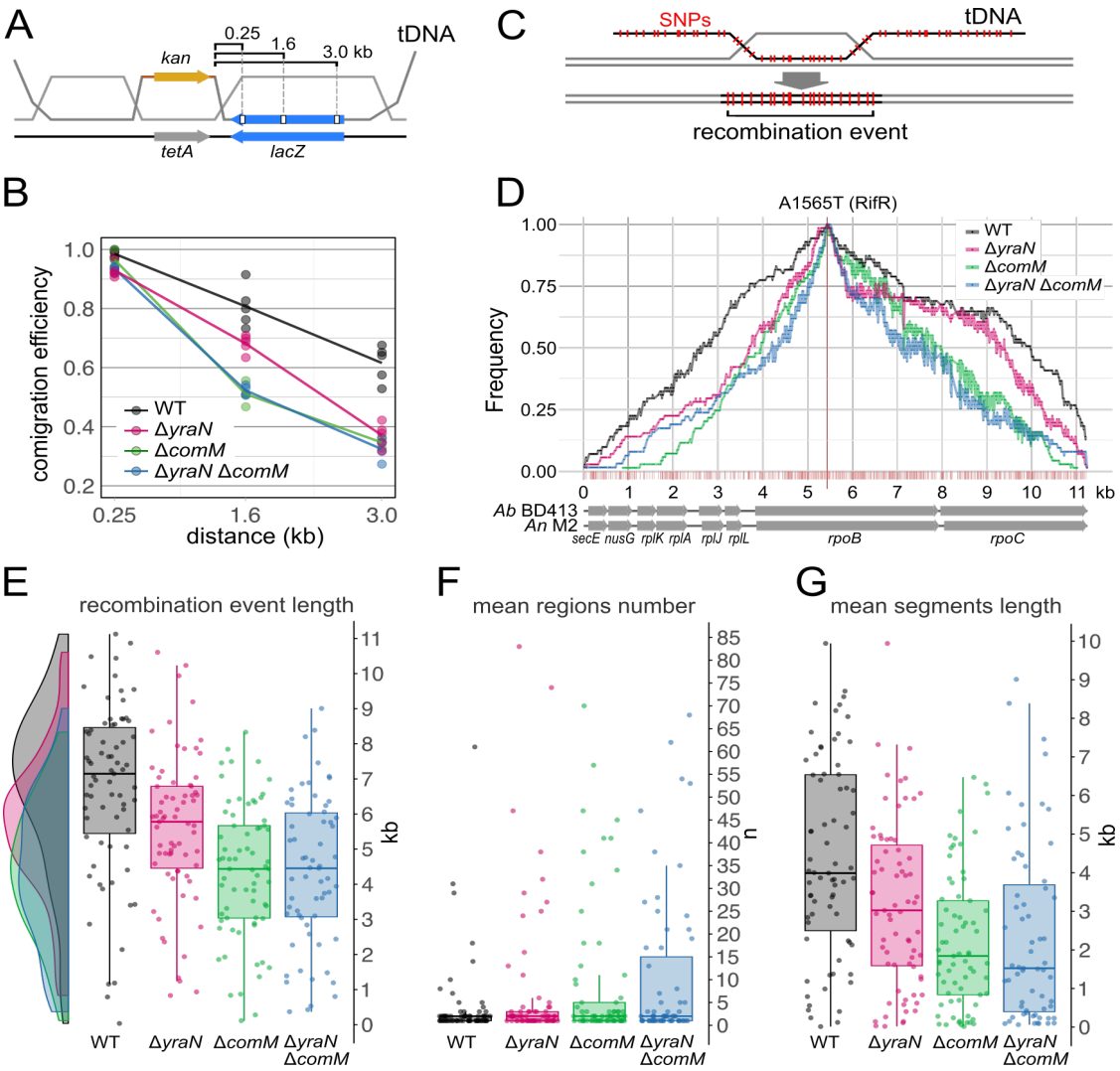

Figure 3. YraN promotes integration of transforming DNA in association with the ComM helicase.
A. Principle of the *in vivo* branch migration assay based on marker comigration. An *A. nosocomialis* strain carrying an integrated tetracycline resistance gene followed by the *lacZ* gene is exposed to transforming DNA (tDNA, dark grey)

encompassing the tetracycline resistance, but replaced by a kanamycin resistance gene, and with a 7-bp deletion in the *lacZ* gene at a distance of 0.25, 1.6 or 3 kb from the end of the kanamycin marker. B. Comigration efficiency (kan^R $lacZ^-$ CFUs vs all kan^R CFUs) as a function of the distance separating the markers in WT and mutants of *A. nosocomialis*. C. Schematic representation of the assay to detect recombination events. *A. baylyi* is exposed to an 11.2-kb PCR fragment from the *rpoB* region and amplified from *A. nosocomialis* resistant to rifamipicin. Over 1,000 SNPs differentiate the sequence from *A. baylyi* and *A. nosocomialis* (red vertical segments). Site-specific sequencing of the transformants is used to determine the acquired SNPs, defining recombination events. D. Frequencies of acquisition of each SNP in the wild-type and mutants of *yraN*, *comM,* or *yraN* and *comM*, averaged from 63 to 71 transformants, obtained in three independent experiments (on different days). The SNP A1565T confers resistance to rifampicin and its acquisition was selected. E. Quantitative analyses of the recombination events observed in panel D. The lengths of recombination events is determined as the distance separating the distal SNPs acquired in each transformants. F. Average number of recombined regions per transformant, defined by one or more acquired sequential SNP. G. Average length per transformant of recombined segments, defined as regions of more than one acquired sequential SNP. Wiskers delineate the range, boxes delineate the 1st and 3rd quartiles, horizontal line shows the median.

Discussion

We here report on the identification of YraN as a core player in the pathway of natural transformation. The function of YraN was, until now, unknown. Using comparative Tn-seq analyses in two distinct Gram-negative bacteria, *L. pneumophila* and *S. elongatus,* we confirmed that the transformation pathway requires genes encoding the components of the Type IV pilus, and all the core genes previously identified as being specifically involved in natural transformation. Only a single other gene was required for transformation, *yraN*. This suggests that transformation does not involve any other conserved gene that is non-essential for viability. Being a member of a large family of phosphodiesterase, one plausible role for YraN was to process the periplasmic dsDNA into the ssDNA as it crosses the inner membrane. However, the *yraN* mutant is not defective in transformation of a plasmid that can replicate in the cytoplasm (Figure S1A). This readily excluded a generic role in DNA uptake. This, in turn, and given the list of genes required for transformation in *L. pneumophila* and *S. elongatus*, also indicates that the nuclease activity involved in the conversion of dsDNA to ssDNA is either carried out by one of the known components of the pathway (ComEC, ComFC) or by an enzyme that is either distinct between species or essential for viability. In addition to dispensability for transformation of replicating plasmids, co-occurrence and co-evolution analyses pointed to a specific involvement of YraN in the integration of transforming DNA. We indeed confirmed that YraN could interact with ComM (Figure 1), that together contribute to increase transformation rates (Figure 2), and analyses of recombination events produced by mutants of *yraN* and *comM* indicate that both are needed to produce extended recombination events (Figure 3). While the helicase activity of ComM in extending the transformation D-loop can explain higher transformation rates and the production of larger recombination events, the contribution of a nuclease was unexpected.

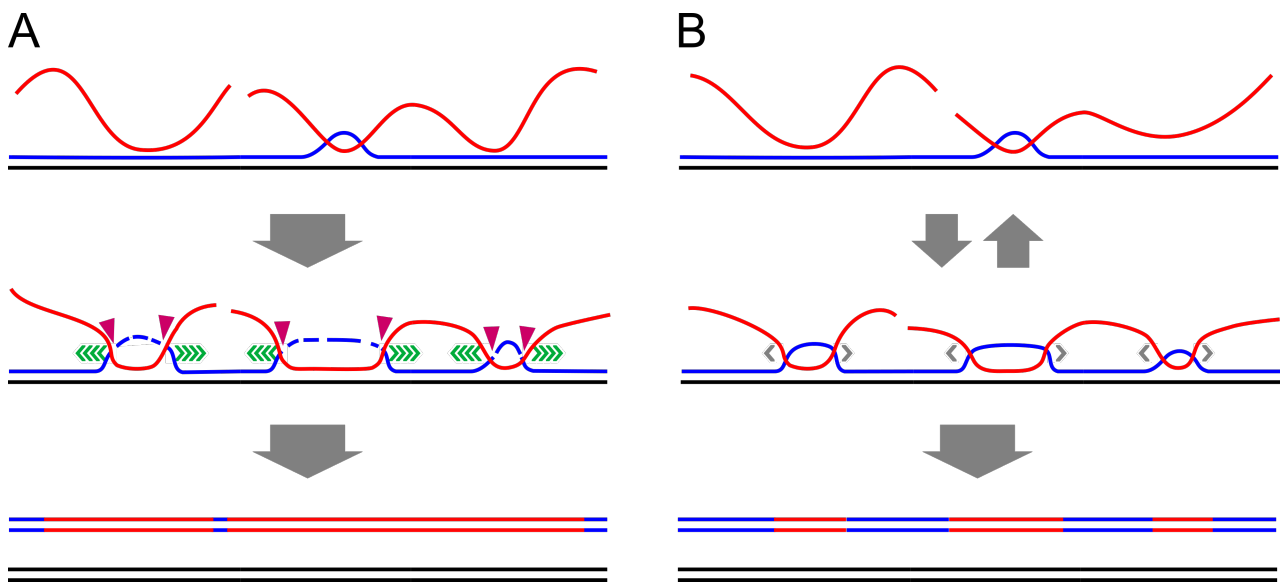

Figure 4. Model of transforming DNA integration involving the YraN-ComM nuclease-helicase system.
A. Model involving the YraN-ComM system in diderm bacteria. Transforming DNA (homologous or homeologous), imported as ssDNA (red line) is covered by DprA and RecA promotes homology-mediated strand invasion. This causes the displacement of the non-complementary strand (blue) and the formation of a D-loop. Loading of ComM stimulates the bidirectional extension of the D-loop (green arrows) while YraN, recruited by its interaction with ComM, cleaves the displaced strand (pink downward-facing triangles). Cleavage of the displaced strand forbids reversal of the D-loop and relieves any topological stress that could limit ComM-mediated D-loop extension. Several strand invasions, from one or more DNA molecules, form as many D-loops, which could merge and, and following replication, create a seemingly continuous recombination event in one chromosome and an original region in the other sister chromosome. B. In the absence of YraN, ComM, or both, D-loops form as efficiently, but might be reversed at higher rates, lowering transformation frequencies. D-loop might be extended by a less processive alternate pathway, leading to shorter and more fragmented recombination events.

How could cleavage of a ssDNA by YraN promote integration of transforming DNA? The mechanisms of recombination of transforming ssDNA are poorly understood and might differ from other recombination pathways. We propose a model for the integration of transforming DNA involving YraN, as depicted in Figure 4. This ssDNA (highlighted in red, Figure 4) is loaded with DprA and RecA, that shields it from nuclease activity (33). Strand invasion mediated by RecA displaces one chromosomal strand (blue strand) to allow for the formation of a heteroduplex between the tDNA and the other chromosomal strand (black strand, complementary to the tDNA). This creates a D-loop that recruits the hexameric ComM helicase at the recombination site, which then unwinds the chromosomal DNA and promotes the extension of the D-loop (green arrows, Figure 4A) (16). Only about 25% of invasion events followed by the recruitment of ComM successfully produced recombinants (16), suggesting that strand invasion and initial D-loop extension are reversible events. Through its interactions with ComM and DprA, YraN is also recruited at the D-loop to introduce one or more cleavages in the displaced strand (pink diamonds, Figure 4A). This would prevent the reversion of strand displacement, stabilizing a now-opened D-loop, and resulting in a higher success rate of recombination events. It is unclear how YraN would specifically cleave the displaced strand. This cleavage might not be (or loosely) sequence-specific

(as suggested by Figure 1) and simply relied on the displaced strand being more exposed to nucleases (as it is not involved in a duplex nor protected by DprA like the tDNA strand). Then, the combined action of ComM and YraN would allow several D-loops to converge, forming continuous recombination events. In their absence, D-loops would still form and extend through an unknown and less processive alternate pathway, resulting in multiple shorter recombination events (Figure 4B).

In addition to preventing D-loop reversal, YraN may ease D-loop extension promoted by ComM. Based on the hypothesis that ComM unwinds chromosomal DNA during D-loop extension, it is anticipated that D-loop extension produces a topological stress which may impede further extension. Cleavage of the displaced strand by YraN, acting as a relaxase, would relieve the topological stress, ensuring that ComM extends the D-loop over long distances (several kb). This possible role of YraN is consistent with the fact that the *yraN* mutant shows weaker comigration defect than the *comM* mutant at short distances (1.6 kb), while the two mutants display stronger and identical defect at longer distances (3 kb) (Figure 3B). In this model, YraN is positioned at the back of ComM, where the interaction of YraN with DprA directs it toward cleavage of the displaced strand. Yet, how ComM is loaded and positioned at the D-loop is not known, and further work is needed to establish the structural configuration of this nuclease/helicase complex. Of note, the contribution of YraN in forming extended recombination events retrospectively explains the initial observation of a 3-times stronger transformation defect of the *yraN* mutants in *A. baylyi* than in *L. pneumophila* (120-fold reduction in *Ab* vs 40-fold in *Lp*) (Figure 1). Transformation in *L. pneumophila* was tested with a 4-kb long tDNA with a centered selectable SNP while in *A. baylyi* the selectable SNP is centered on 11 kb-long tDNA. This is a relevant difference since the importance of YraN is expected to increase with the length of the region surrounding the selected SNP: the longer this region is, the more chances there are for the initial invasion event to occur far from the selected SNP, therefore relying on efficient branch migration to be converted.

The analysis of recombination events, based on the detection of SNPs acquisition, may also indicate an alternative and/or complementary role of YraN/ComM to D-loop extension, specifically when incorporating homeologous DNA. Some recombination events, display extensively fragmented structures (number of regions>10, Figure 3E), most often found at one edge of recombination events (Figure S5). These highly fragmented events are more frequent in the *comM* and *comM yraN* mutants (n=14 and n=17, respectively) than in the wild-type (n=5) (Figure 3E). A plausible explanation is that ComM-YraN complexes protect homeologous DNA from the mismatch repair (MMR) pathway. MMR, which would initially recognize a transition (34), could also resolve surrounding mismatches (transitions or transvertions) creating mosaicism in the recombination

event. Shielding the heteroduplex from MMR implies that multiple ComM-YraN complexes could cover the heteroduplex, in addition to acting at the edges of the D-loop.

Our work suggests that the ComM-YraN helicase/nuclease system evolved to favor integration of large, potentially polymorphic sequences. Interestingly, the phylogenetic distribution of this system is consistent with differences in the length of recombination events produced by transformation. Genome-wide analyses of recombination events in transformable diderms species have found that natural transformation produces large recombination events, with a mean length of 10 to 14 kb in *Haemophilus influenzae* (35) and *Neisseria meningitidis* (36). In addition, under mixed-culture conditions, where DNA is naturally released, the mean length of the total acquired DNA in *V. cholerae* is even greater and was found to be of ~70 kb (37). In contrast, in the monoderms *B. subtilis* and *S. pneumoniae*, the mean length of recombination events is lower, even in mixed-culture setup, ranging from 2 to 4 kb (38–40). These monoderms and most Firmicutes lack YraN and ComM (Figure 2). Indeed, recent studies in *S. pneumoniae* and *B. subtilis* propose that firmicutes, instead of employing a specialized helicase, rely on the effector of homologous recombination RadA to promote extension of the transformation D-loop (41, 42). We speculate that such helicase may not be as processive as the ComM-YraN complex.

In all, the identification of YraN as a dedicated nuclease involved in promoting large recombination events sheds light on the specific involvement of a helicase/nuclease complex. Further investigation is needed to unravel its contribution to the still enigmatic mechanisms underlying the integration of DNA through natural transformation. Yet, this broadly conserved helicase/nuclease system is another indication that most bacteria possess a complete and dedicated pathway for the transport and recombination of exogenous DNA.

Material and methods

Bacteria and growth conditions

All the *Legionella pneumophila* strains are derived from the Paris or Lp02 strains. They were cultured in solid medium containing ACES [*N*-(2-acetamido)-2-aminoethanesulfonic acid] and buffered charcoal yeast extract (CYE) or in liquid medium containing ACES and buffered yeast extract (AYE). When appropriate, media were supplemented with kanamycin (15 µg/mL for Paris strain and 50 µg/mL for Lp02 strains), streptomycin (50 µg/mL), gentamycin (10 µg/mL) or apramycin (15 µg/mL). All the *Acinetobacter* strains are derived from *A. nosocomialis* M2 or *A. baylyi* BD413 (also known as ADP1). When appropriate, they were grown in LB broth or LB agar supplemented with tetracycline (5 µg/mL), kanamycin (50 µg/mL), apramycin (30 µg/mL), or

rifampicin (100 µg/mL). The methods for constructing mutants derived from these different strains were previously reported (43–45) and described in the Supplementary material.

Transformation assays

For transformation assays of *L. pneumophila*, the parent strain (denoted wild-type, WT) is a constitutively transformable strain obtained by introducing a premature stop codon in the post-transcriptional repressor of competence *rocC* (43, 44). Overnight streak of the desired *L. pneumophila* Paris strain was resuspended in 1 mL sterile water to an OD600 of 1 to obtain ~1.10^9 CFU/mL. A 10-µL portion of this suspension (~1.10^7 CFUs) was spotted on CYE plate supplemented with 0.5 mM IPTG when appropriate. Then, 10 µL of transforming DNA (a 4 kb-long PCR product encompassing the *rpsL* mutation conferring resistance to streptomycin (46)) at the indicated concentration was spotted directly on the bacteria spot and left absorbed onto the agar. Plates were incubated at 37°C for 24 hours. Each spot was then resuspended in 200 µL sterile water and ten-fold serial dilutions were plated on non-selective and selective medium. Plates were incubated at 37°C for 72 hours and CFU counting was performed.

Transformation assays of *A. baylyi* were conducted as follows: 70 µL of overnight culture in LB liquid medium at 30°C of the desired strain was used to inoculate 1 mL LB. After 1 hour incubation at 30°C with shaking, tDNA was added at the indicated concentration and the cultures were re-incubated at 30°C with shaking for 1 hour to overnight and followed by DNAse I treatment to stop DNA uptake. Ten-fold serial dilutions were then plated on non-selective or selective medium. Plates were incubated at 30°C overnight and CFU counting was performed. Transformation assays of *A. nosocomialis* were performed as previously described (45, 47). 2 mL LB were inoculated with overnight streak of the desired strain and incubated at 37°C until reaching an OD600 of 1 to obtain ~1.10^9 CFU/mL. The culture was 100-fold diluted in sterile water to obtain 10^7 CFU/mL. Then, 10 µL of this suspension was mixed with 10 µL of transforming DNA at the indicated concentration and 2.5 µL of this mixture was stabbed into freshly prepared motility medium (5 g/liter agarose D3 (Euromedex), 5 g/liter tryptone, and 2.5 g/liter NaCl) in 2-mL Eppendorf tubes. Tubes were incubated at 37°C overnight and after incubation time, bacteria were resuspended in 200 µL sterile water by vigorous vortexing. Ten-fold serial dilutions of this suspension were then plated on non-selective or selective medium. Plates were incubated at 37°C overnight and CFU counting was performed. Transformation frequencies were calculated as the ratio of the number of CFU counted on selective medium (i,e,, the transformants) to the number of CFU counted on non-selective medium (i,e,, total population). All transformation assays were performed at least three times independently (on different days).

Phylogenetic analyses

21,105 complete bacterial genomes present in RefSeq were downloaded in March 2021. The dataset was searched for the following hmm profiles (hmmsearch from hmmer v.3.3.2) (48): PF02021.20 for *yraN*, the co-occurrence of PF13541 (chlI domain), PF01078 (MgChelatase) and PF1335 (MgChelatase C domain) for *comM*, PF00154.24 for *recA*, PF02481.18 for *dprA*. A protein was considered present in the genome when the sequence score of the alignment was above the gathering threshold score cutoff (--cut_ga) and covered more than 50% of the hmm profile. Several proteins matched both ComM and YraN domains (in different positions). Since these genomes lacked hits of both proteins, we interpreted this as a fusion of YraN and ComM. The Genome Taxonomy Database (GTDB) (49) (March 2022, version r207) was used to pick representative strains for every bacterial species and build a bacterial phylogenetic tree. Among the 21,105 bacterial genomes of the dataset, 4,185 strains were represented in GTDB tree. This subsampling retained ca. 80% of the genera initially present. This subset of the GTDB tree obtained with the get_subtree_withtips function from castor R package was used to draw a tree. The phylum annotations of the strains were provided by GTDB. The Firmicutes, Desulfobacterota and Nitrospirota phyla were each divided into several phyla by GTDB. They were grouped together to have a unique phylum for each of them and simplify the reading of the tree. Two proteins were considered co-localized in a genome when they are at less than 10 genes of distance. The statistical significance of the co-occurrence of genes in a genome was corrected for the phylogenetic structure of the bacterial population. A phylogenetic logistic regression (phyloglm function from phylolm R package (50)) was used, to which we provided the GTDB phylogenetic tree. The data were fitted using the 'logistic_MPLE" method (btol=20, boot=100).

Bacterial two-hybrid system experiments

Two hybrid experiments were carried out to test the interaction between proteins involved in transformation-associated homologous recombination. To that end, *recA*, *dprA*, *yraN* and *comM* genes from *L. pneumophila* Paris were cloned into pKT25, pKNT25, pUT18, or pUT18C plasmids in frame with T25 or T18 catalytic domains of *B. pertussis* adenylate cyclase (32) by Sequence and Ligation Independent Cloning (SLIC) (51) (see supplementary informations for more details). With this method, for each protein of interest, 2 recombinant plasmids encoding 2 hybrid proteins were propagated in their respective *Escherichia coli* DH5a competent hosts: pKNT25-X and pUT18-X, encoding respectively X-T25 and T18-X. Each plasmid encoding an X-T25 fusion protein was then co-transformed in competent BTH101 *cya-* strains with a plasmid encoding X-T18 fusion protein. The co-transformed bacteria obtained were pre-cultured in LB liquid medium supplemented with 25 µg/mL kanamycin and 100 µg/mL ampicillin at 37°C overnight with shaking. After incubation,

these cultures were 1:10 diluted and spotted onto LB agar supplemented with 25 µg/mL kanamycin, 100 µg/mL ampicillin, 0.5 mM IPTG and 100 µg/mL X-β-Gal (Roth) and incubated at 20°C for 48 to 72 hours. Interaction between DprA and RecA are used as positive controls in this experiment since their interaction was already described (14).

Expression and purification of YraN

In order to produce C-terminal His tagged YraN in *E. coli*, *yraN* from *L. pneumophila* was amplified by PCR and introduced in pKHS vector (a pET28 derivative) by InFusion methodology. The E50A variant was further obtained by site-directed mutagenesis using an adapted protocol from the Stratagene "QuickChange" with only one mutator primer carrying the modified sequence and with both the Pfu turbo polymerase (Stratagene) and the Taq ligase (New England Biolabs) in the PCR reaction. pKHS-yraN and its derivative pKHS-yraN-E50A were introduced in *E. coli* BL21-Rosetta(DE3)-pLysS (CmR) cells (Novagen). The expression clones were grown at 30°C in 400 ml LB medium with Cm (10 µg/ml), Kan (50 µg/ml), and 0.2% glucose until the OD600 reached 0.7. Protein expression was then induced by adding 0.5 mM IPTG followed by growth at 30°C for 2 hours. Cell pellets were collected by centrifugation (6000 g; 20 min; 4°C) and further resuspended in 30 ml solution containing 50 mM Na-Phosphate pH 7.5, 300 mM NaCl, 10 mM imidazole before breaking cells and DNA with sonifier (Vibracell). The soluble extracts obtained after clarification centrifugation (15000 g; 30 min; 4°C in JA20) were injected first onto a His-Trap crude 1 ml column and further purified by FPLC (Äkta purifier-10, GE Healthcare) using a solution of 50 mM Na-Phosphate pH 7.5, 300 mM NaCl, and an increasing gradient of imidazole. The eluted fractions were diluted 2.5 fold in 50 mM Na-Phosphate pH 7.5, to reach 120 mM of NaCl and were further passed through a Hi-Trap Q HP 1 mL column to remove residual DNA before binding on a Hi-Trap heparin HP 1 mL column (GE Healthcare) and elution with a NaCl gradient. Purified YraN proteins or variants were finally injected and separated on a Superdex 75 16/600 sizing column (GE healthcare) equilibrated with 50 mM Na-Phosphate pH 7.5, 200 mM NaCl. Part of the purified protein was stored at -80°C after addition of 10% glycerol, 1 mM TCEP for longer storage or at -20°C after addition of 50% glycerol. Another part was dialyzed against 50 mM Hepes pH 7.6, 200 mM NaCl, 0,5 mM TCEP, 10% glycerol using PD miniTraP 25 buffer exchange column for buffer comparative assays on YraN activities. The purity and the concentrations of the purified proteins were analyzed by SDS-PAGE and Coomassie staining analysis.

Nuclease assay

Increasing amounts of YraN and YraN-E50A (150 to 1,000 nM) were incubated 30 min at 37°C with either 100 ng of supercoiled pUC18 plasmid or 10 nM of Cy3-labeled DNA substrates, in 10

µL solutions containing 50 mM Hepes pH 7.6, 50 mM NaCl, 0,1 mg/ml BSA, 0,5 mM TCEP, 6% glycerol, 0,3 to 3 mM $MnCl_2$ or/and 10 mM $MgO(acetate)_2$. In order to quench the reaction and to deproteinize the DNA, 0.1% SDS and 1 mM EDTA were added to the reaction for 5 min at 37°C before loading half of the reaction on 1% agarose gel / TAE 1X for long DNA molecules or on 4% acrylamide (29:1) / 0.3X TBE gels for short DNA structures (100 nt). Cy3 labeled oligonucleotides were detected with the Typhoon Trio imager and the cold plasmid DNA was detected after ethidium bromide staining.

Comigration marker experiment

Comigration marker experiment was adapted from the comigration assay conducted in *V. cholerae* (17). This experiment was conducted in *A. nosocomialis* M2 as follows. Strains that carry genetically linked markers: *lacZ* gene placed adjacent to a tetracycline cassette. Details of strain construction can be found in the Supplementary material. This strain called "$lacZ^+$-tet" was naturally transformed with the following PCR products: "2kB-$lacZ^{\mathbf{dlt1}}$-kan-2kB" or "2kB-$lacZ^{\mathbf{dlt2}}$-kan-2kB" or "2kB-$lacZ^{dlt3}$-kan-2kB" as tDNA. These tDNA have 2 kB upstream and downstream homologous regions with "$lacZ^+$-tet" recipient strains that allow the acquisition of the tetracycline cassette (replacing the kanamycin cassette) and of the 7-bp deletion in *lacZ* ($lacZ^{dltX}$) by homologous recombination. The 7 pb deletions were designed so that (i) they are not substrates for the mismatch repair and (ii) they shift the reading frame of *lacZ* to produce a TGA stop codon resulting in *lacZ*- phenotype. These 7-bp deletions were designed at increasing distances from the original TAA stop codon of the kanamycin cassette (distance between kan_{TAA} and $lacZ_{TGA}$ for dlt1=3,043bp, for dlt2=253bp, and for dlt3=1,606bp). Transformation assays were conducted as described above, mixing a 5-µL suspension of each mutant with 5 µL of purified PCR (150 ng/µl) dlt1, dlt2, or dlt3, followed by incubation at 37°C for 24 hours. Ten-fold serial dilutions of the transformation cultures were plated on LB agar supplemented with 50 µg/mL kanamycin and 100 µg/mL X-gal or on nonselective LB plates and incubated in appropriate conditions described before. Comigration efficiency, which refers to the number of kanamycin-resistant transformants that have acquired the *lacZ* deletion, was calculated as the ratio of the number of white transformants (*kanR/lacZ*$^-$) divided by the number of total transformants (*KanR/lacZ*$^+$ + *KanR/lacZ*$^-$) for each tDNA and each mutant.

Characterization of recombination events

Direct detection of transformation recombination tracts was carried out as follows. *A. baylyi* strain was transformed with an 11.2 kb-long PCR product as tDNA that confers resistance to rifampicin upon recombination. This tDNA was obtained by amplifying an 11.2-kb portion of *A. nosocomialis*

M2 rif^R genome (52) that carries the *rpoB* gene, using forward and reverse primers hybridizing at nearly equal distance from the mutation conferring resistance to rifampicin (see primer list in the Supplementary material). *A. baylyi* and mutants were transformed with 100 ng of DNA for 1 hour, DNAse I was added to the culture, further incubated for 1 hour, and plated on rifampicin-containing plates. In each of the three independent experiments, 24 colonies were picked for each tested strain and the region of interest was amplified as two overlapping PCR products, which were used to construct a sequencing library (Rapid Barcoding Kit 96 SQK-RBK110.96, Oxford Nanopore) and sequenced using a MinIon instrument (Oxford Nanopore). Reads were basecalled using the super-accuracy model (sup, guppy 5) and filtered on coverage. Variant calling was performed using Medaka (v1.3.4) using the donor and recipient sequence as reference, and filtered on quality. VCF files were processed with R (4.3.2 Eye Holes) in order to identify converted SNPs, extract recombination events and regions, and perform statistical analysis on obtained distributions. Converted SNP corresponds to haplotype which differs between donor and recipient, and match donor haplotype. Recombination events are defined by the two outermost SNPs of continuous or discontinuous series of acquired SNPs. Regions are defined as strictly continuous series of SNPs and isolated SNPs.

Acknowledgments

We thank Marie Cariou (CIRI) for her assistance in writing scripts for analyses of preliminary experiment on recombination events and Clothilde Rousseau for proofreading the manuscript. This work was supported in part by grant ANR-20-CE12-0004 (TransfoConflict) from Agence Nationale de la Recherche (awarded to XC and EPCR). LH is supported by a fellowship "Fin de thèse" from the Fondation pour la Recherche Médicale (FDT202001010890). This work used the computational and storage services (TARS cluster) provided by the IT department at Institut Pasteur, Paris.

References

1. Johnston C, Martin B, Fichant G, Polard P, Claverys J-P. 2014. Bacterial transformation: distribution, shared mechanisms and divergent control. Nat Rev Microbiol 12:181–196.
2. Dubnau D, Blokesch M. 2019. Mechanisms of DNA Uptake by Naturally Competent Bacteria. Annu Rev Genet 53:217–237.
3. Laurenceau R, Péhau-Arnaudet G, Baconnais S, Gault J, Malosse C, Dujeancourt A, Campo N, Chamot-Rooke J, Le Cam E, Claverys J-P, Fronzes R. 2013. A Type IV Pilus Mediates DNA Binding during Natural Transformation in Streptococcus pneumoniae. PLoS Pathog 9:e1003473.

4. Ellison CK, Dalia TN, Vidal Ceballos A, Wang JC-Y, Biais N, Brun YV, Dalia AB. 2018. Retraction of DNA-bound type IV competence pili initiates DNA uptake during natural transformation in Vibrio cholerae. Nat Microbiol 3:773–780.

5. Berry J-L, Pelicic V. 2015. Exceptionally widespread nanomachines composed of type IV pilins: the prokaryotic Swiss Army knives. FEMS Microbiol Rev 39:134–154.

6. Lam T, Ellison CK, Eddington DT, Brun YV, Dalia AB, Morrison DA. 2021. Competence pili in Streptococcus pneumoniae are highly dynamic structures that retract to promote DNA uptake. Mol Microbiol 116:381–396.

7. Zuke JD, Erickson R, Hummels KR, Burton BM. Visualizing dynamic competence pili and DNA capture throughout the long axis of Bacillus subtilis. J Bacteriol 205:e00156-23.

8. Provvedi R, Dubnau D. 1999. ComEA is a DNA receptor for transformation of competent *Bacillus subtilis*. Molecular Microbiology 31:271–280.

9. Seitz P, Pezeshgi Modarres H, Borgeaud S, Bulushev RD, Steinbock LJ, Radenovic A, Dal Peraro M, Blokesch M. 2014. ComEA is essential for the transfer of external DNA into the periplasm in naturally transformable Vibrio cholerae cells. PLoS Genet 10:e1004066.

10. Draskovic I, Dubnau D. 2005. Biogenesis of a putative channel protein, ComEC, required for DNA uptake: membrane topology, oligomerization and formation of disulphide bonds. Mol Microbiol 55:881–896.

11. Diallo A, Foster HR, Gromek KA, Perry TN, Dujeancourt A, Krasteva PV, Gubellini F, Falbel TG, Burton BM, Fronzes R. 2017. Bacterial transformation: ComFA is a DNA-dependent ATPase that forms complexes with ComFC and DprA. Molecular Microbiology 105:741–754.

12. Sinha S, Mell JC, Redfield RJ. 2012. Seventeen Sxy-Dependent Cyclic AMP Receptor Protein Site-Regulated Genes Are Needed for Natural Transformation in Haemophilus influenzae. J Bacteriol 194:5245–5254.

13. Seitz P, Blokesch M. 2013. DNA-uptake machinery of naturally competent Vibrio cholerae. PNAS 110:17987–17992.

14. Mortier-Barrière I, Velten M, Dupaigne P, Mirouze N, Piétrement O, McGovern S, Fichant G, Martin B, Noirot P, Le Cam E, Polard P, Claverys J-P. 2007. A key presynaptic role in transformation for a widespread bacterial protein: DprA conveys incoming ssDNA to RecA. Cell 130:824–36.

15. Bergé M, Mortier-Barrière I, Martin B, Claverys J-P. 2003. Transformation of Streptococcus pneumoniae relies on DprA- and RecA-dependent protection of incoming DNA single strands. Mol Microbiol 50:527–536.

16. Dalia AB, Dalia TN. 2019. Spatiotemporal Analysis of DNA Integration during Natural Transformation Reveals a Mode of Nongenetic Inheritance in Bacteria. Cell 179:1499-1511.e10.

17. Nero TM, Dalia TN, Wang JC-Y, Kysela DT, Bochman ML, Dalia AB. 2018. ComM is a hexameric helicase that promotes branch migration during natural transformation in diverse Gram-negative species. Nucleic Acids Res 46:6099–6111.

18. Albano M, Hahn J, Dubnau D. 1987. Expression of competence genes in Bacillus subtilis. J Bacteriol 169:3110–3117.

19. Fani R, Mastromei G, Polsinelli M, Venema G. 1984. Isolation and characterization of Bacillus subtilis mutants altered in competence. J Bacteriol 157:152–157.

20. Gwinn ML, Ramanathan R, Smith HO, Tomb JF. 1998. A new transformation-deficient mutant of Haemophilus influenzae Rd with normal DNA uptake. J Bacteriol 180:746–748.

21. Karudapuram S, Zhao X, Barcak GJ. 1995. DNA sequence and characterization of Haemophilus influenzae dprA+, a gene required for chromosomal but not plasmid DNA transformation. J Bacteriol 177:3235–3240.

22. Hardy L, Juan P-A, Coupat-Goutaland B, Charpentier X. 2021. Transposon Insertion Sequencing in a Clinical Isolate of Legionella pneumophila Identifies Essential Genes and Determinants of Natural Transformation. J Bacteriol 203:e00548-20.

23. Taton A, Erikson C, Yang Y, Rubin BE, Rifkin SA, Golden JW, Golden SS. 2020. The circadian clock and darkness control natural competence in cyanobacteria. Nat Commun 11:1688.

24. Huerta-Cepas J, Forslund K, Coelho LP, Szklarczyk D, Jensen LJ, von Mering C, Bork P. 2017. Fast Genome-Wide Functional Annotation through Orthology Assignment by eggNOG-Mapper. Mol Biol Evol 34:2115–2122.

25. Laganeckas M, Margelevicius M, Venclovas C. 2011. Identification of new homologs of PD-(D/E)XK nucleases by support vector machines trained on data derived from profile-profile alignments. Nucleic Acids Res 39:1187–1196.

26. Steczkiewicz K, Muszewska A, Knizewski L, Rychlewski L, Ginalski K. 2012. Sequence, structure and functional diversity of PD-(D/E)XK phosphodiesterase superfamily. Nucleic Acids Research 40:7016–7045.

27. Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, Tunyasuvunakool K, Bates R, Žídek A, Potapenko A, Bridgland A, Meyer C, Kohl SAA, Ballard AJ, Cowie A, Romera-Paredes B, Nikolov S, Jain R, Adler J, Back T, Petersen S, Reiman D, Clancy E, Zielinski M, Steinegger M, Pacholska M, Berghammer T, Bodenstein S, Silver D, Vinyals O, Senior AW, Kavukcuoglu K, Kohli P, Hassabis D. 2021. Highly accurate protein structure prediction with AlphaFold. Nature 596:583–589.

28. Christensen LM, Sule P, Strain M, Cirillo JD. 2019. Legionella pneumophila p45 element influences host cell entry and sensitivity to sodium. PLoS One 14:e0218941.

29. Moran NA, Mira A. 2001. The process of genome shrinkage in the obligate symbiont Buchnera aphidicola. Genome Biology 2:research0054.1.

30. Cong Q, Anishchenko I, Ovchinnikov S, Baker D. 2019. Protein interaction networks revealed by proteome coevolution. Science.

31. A multi-scale coevolutionary approach to predict interactions between protein domains. https://journals-plos-org.proxy.insermbiblio.inist.fr/ploscompbiol/article?id=10.1371/journal.pcbi.1006891. Retrieved 1 September 2021.

32. Karimova G, Pidoux J, Ullmann A, Ladant D. 1998. A bacterial two-hybrid system based on a reconstituted signal transduction pathway. PNAS 95:5752–5756.

33. Bergé M, Mortier-Barrière I, Martin B, Claverys J-P. 2003. Transformation of Streptococcus pneumoniae relies on DprA- and RecA-dependent protection of incoming DNA single strands. Mol Microbiol 50:527–536.

34. Young DM, Ornston LN. 2001. Functions of the Mismatch Repair Gene mutS from Acinetobacter sp. Strain ADP1. J Bacteriol 183:6822–6831.

35. Mell JC, Shumilina S, Hall IM, Redfield RJ. 2011. Transformation of natural genetic variation into Haemophilus influenzae genomes. PLoS Pathog 7:e1002151.

36. Alfsnes K, Frye SA, Eriksson J, Eldholm V, Brynildsrud OB, Bohlin J, Harrison OB, Hood DW, Maiden MCJ, Tønjum T, Ambur OH. 2018. A genomic view of experimental intraspecies and interspecies transformation of a rifampicin-resistance allele into Neisseria meningitidis. Microb Genom 4.

37. Matthey N, Stutzmann S, Stoudmann C, Guex N, Iseli C, Blokesch M. 2019. Neighbor predation linked to natural competence fosters the transfer of large genomic regions in Vibrio cholerae. eLife 8:e48212.

38. Power JJ, Pinheiro F, Pompei S, Kovacova V, Yüksel M, Rathmann I, Förster M, Lässig M, Maier B. 2021. Adaptive evolution of hybrid bacteria by horizontal gene transfer. PNAS 118.

39. Cowley LA, Petersen FC, Junges R, Jimenez MJD, Morrison DA, Hanage WP. 2018. Evolution via recombination: Cell-to-cell contact facilitates larger recombination events in Streptococcus pneumoniae. PLOS Genetics 14:e1007410.

40. Croucher NJ, Harris SR, Barquist L, Parkhill J, Bentley SD. 2012. A High-Resolution View of Genome-Wide Pneumococcal Transformation. PLOS Pathogens 8:e1002745.

41. Marie L, Rapisarda C, Morales V, Bergé M, Perry T, Soulet A-L, Gruget C, Remaut H, Fronzes R, Polard P. 2017. Bacterial RadA is a DnaB-type helicase interacting with RecA to promote bidirectional D-loop extension. Nat Commun 8:15638.

42. Torres R, Serrano E, Alonso JC. 2019. Bacillus subtilis RecA interacts with and loads RadA/Sms to unwind recombination intermediates during natural chromosomal transformation. Nucleic Acids Res 47:9198–9215.

43. Attaiech L, Boughammoura A, Brochier-Armanet C, Allatif O, Peillard-Fiorente F, Edwards RA, Omar AR, MacMillan AM, Glover M, Charpentier X. 2016. Silencing of natural transformation by an RNA chaperone and a multitarget small RNA. Proc Natl Acad Sci USA 113:8813–8818.

44. Juan P-A, Attaiech L, Charpentier X. 2015. Natural transformation occurs independently of the essential actin-like MreB cytoskeleton in Legionella pneumophila. Scientific Reports 5:16033.

45. Godeux A-S, Svedholm E, Lupo A, Haenni M, Venner S, Laaberki M-H, Charpentier X. 2020. Scarless Removal of Large Resistance Island AbaR Results in Antibiotic Susceptibility and Increased Natural Transformability in Acinetobacter baumannii. Antimicrob Agents Chemother 64.

46. Durieux I, Ginevra C, Attaiech L, Picq K, Juan P-A, Jarraud S, Charpentier X. 2019. Diverse conjugative elements silence natural transformation in Legionella species. PNAS 116:18613–18618.

47. Godeux A-S, Lupo A, Haenni M, Guette-Marquet S, Wilharm G, Laaberki M-H, Charpentier X. 2018. Fluorescence-Based Detection of Natural Transformation in Drug-Resistant Acinetobacter baumannii. Journal of Bacteriology 200:e00181-18.

48. Eddy SR. 2011. Accelerated Profile HMM Searches. PLOS Computational Biology 7:e1002195.

49. Parks DH, Chuvochina M, Chaumeil P-A, Rinke C, Mussig AJ, Hugenholtz P. 2020. A complete domain-to-species taxonomy for Bacteria and Archaea. 9. Nat Biotechnol 38:1079–1086.

50. Ho L si T, Ané C. 2014. A linear-time algorithm for Gaussian and non-Gaussian trait evolution models. Syst Biol 63:397–408.

51. Li MZ, Elledge SJ. 2012. SLIC: a method for sequence- and ligation-independent cloning. Methods Mol Biol 852:51–59.

52. Godeux A-S, Sveldhom E, Barreto S, Potron A, Venner S, Charpentier X, Laaberki M-H. 2021. Interbacterial transfer of carbapenem resistance and large antibiotic resistance islands by natural transformation in pathogenic Acinetobacter.