Discovery of the widespread site-specific single-stranded nuclease family Ssn

Martin Chenal [1,4], Alex Rivera-Millot [1,4], Luke B. Harrison [1], Ahmed S. Khairalla[1,3], Cecilia Nieves [1], Ève Bernet [1], Mansoore Esmaili[1], Manel Belkhir[1], Jonathan Perreault [2] & Frédéric J. Veyrier [1] ✉

Site-specific endonucleases that exclusively cut single-stranded DNA have hitherto never been described and constitute a barrier to the development of ssDNA-based technologies. We identify and characterize one such family, from the GIY-YIG superfamily, of widely distributed site-specific single-stranded nucleases (Ssn) exhibiting unique ssDNA cleavage properties. By first comprehensively studying the Ssn homolog from *Neisseria meningitidis*, we demonstrate that it interacts specifically with a sequence (called NTS) present in hundreds of copies and surrounding important genes in pathogenic *Neisseria*. In this species, NTS/Ssn interactions modulate natural transformation and thus constitute an additional mechanism shaping genome dynamics. We further identify thousands of Ssn homologs and demonstrate, in vitro, a range of Ssn nuclease specificities for their corresponding sequence. We demonstrate proofs of concept for applications including ssDNA detection and digestion of ssDNA from RCA. This discovery and its applications set the stage for the development of innovative ssDNA-based molecular tools and technologies.

Nucleases have become an invaluable tool for all aspects of research in molecular biology, from the most basic manipulations to modern gene therapies. These nucleic acid-cutting enzymes are extremely diverse in their specificities, mechanisms, biological functions, and biotechnological uses. Zinc-finger nucleases (ZFN) and Transcription activator-like effector nucleases (TALENs) have programmable specificities used mostly in genetic engineering[1]. The past decade has also seen the development of CRISPR-associated proteins (Cas) into site-specific nucleases capable of cleaving virtually any DNA sequence, unleashing a biotechnological revolution[2]. Among the multiple nuclease families, the GIY-YIG superfamily is composed of endonucleases involved in DNA repair and recombination, restriction of foreign DNA, and mobile genetic elements. They possess a ~100 amino acid nuclease domain forming three antiparallel β-sheets encased in several α-helices, with the signature "GIY" and "YIG" semi-conserved motifs in the catalytic core[3]. GIY-YIG nucleases are typically modular proteins with distinct DNA-binding domains underpinning their evolution into a large phenotypic landscape of structures and sequence specificities[3]. This modularity has been exploited for the generation of chimeric enzymes harboring a GIY-YIG nuclease domain for genome-editing applications[4–6]. However, unlike the modular proteins from this superfamily, single-domain GIY-YIG small proteins such as the cd10448 family remain uncharacterized, with thousands of homologs across the entire bacterial domain.

The genus *Neisseria* is both a remarkable example of bacterial evolution and host adaptation, and a serious public health threat. While most *Neisseria* species are commensals of human and animal mucosal surfaces, two have evolved to become human-restricted pathosymbionts: *Neisseria meningitidis* (*Nm*; meningococcal meningitis and septicemia) and *Neisseria gonorrhoeae* (*Ng*; gonorrhea). The majority of *Neisseria* species are naturally competent, and readily uptake and integrate exogenous DNA into their genomes[7–9]. This trait

[1]INRS-Centre Armand-Frappier Santé Biotechnologie, Bacterial Symbionts Evolution, Laval, QC, Canada. [2]INRS-Centre Armand-Frappier Santé Biotechnologie, Laval, QC, Canada. [3]Present address: Department of Biology, University of Regina, Regina, SK, Canada. [4]These authors contributed equally: Martin Chenal, Alex Rivera-Millot. ✉e-mail: frederic.veyrier@inrs.ca

is associated with a high frequency of horizontal gene transfer[10,11], promoting at once the spread of virulence factors[12–14] and antibiotic resistance genes[15–17] along with genetic diversity mechanisms such as antigenic variation[18]. Unlike most other naturally competent species such as some *Vibrio* spp., *Streptococcus* spp., and *Haemophilus influenzae*, competence in *Neisseria* species is thought to be constitutive[19,20]. No specific regulatory mechanisms have been identified in this genus, and the competent state is maintained throughout growth stages of the bacteria regardless of environmental or intracellular stress such as starvation or DNA damage[21,22]. Interestingly, *Neisseria* exhibits one of the highest transformation frequencies among bacteria, reaching up to 50% in *N. gonorrhoeae*[23]. This frequency was found to be highly variable between species but also between strains of a single species, suggesting the presence of genetic factors involved in the regulation of competence[24,25]. One such factor has been elucidated in *Neisseria* and similarly in *H. influenzae*[26]. These bacteria are able to discriminate and preferentially uptake DNA from closely related species through the recognition of a small repeated sequence, called the DNA Uptake Sequence (DUS), found in thousands of copies in their genomes[27]. In addition to the DUS, another *Neisseria* repeated sequence, the duplicated repeat sequences 3 (dRS3), has been shown to be enriched near genes with interisolate sequence variability, and has been identified as a target for a specific phage recombinase[28–30]. Despite being much more frequent in pathogenic genomes (*Nm* and *Ng*) relative to their commensal counterparts, the association between dRS3 repeats and *Neisseria* competence and virulence is limited to in silico genomic analyses showing their association in the genome with specific families of genes[14,28,31].

In this study, we identified a repeated element of 28nt, termed *Neisseria* Transformation Sequence (NTS), that encompasses the shorter dRS3 sequence. The NTS likely forms a hairpin structure and is an order of magnitude more frequent in the pathogenic *Neisseria* genomes relative to commensal species. We show that it is present in variable numbers in the genomes of selected *Neisseriaceae* and that these sequences cluster non-randomly around certain important genes in *N. meningitidis*. Importantly, we demonstrate here that these repeated NTSs are also spatially associated with a specific gene (*NMV_0044* or *NMB0047* in the *Nm* 8013 and MC58 genomes, respectively) across broad phylogenetic distance in bacteria. These homologs encode small hypothetical proteins and encompass the cd10448 (NCBI Conserved Domain Database): GIY-YIG_unchar_3 protein family and are abundant in Pseudomonadati. We show that the protein from *Nm*, renamed SsnA, is a site-specific single-stranded endonuclease capable of binding and cutting the NTS only as ssDNA. We demonstrate that this ssDNase activity regulates natural competence and modulates the integration of transforming DNA. We also show that SsnA belongs to a vast family of small single-domain endonucleases (Ssn) with unique specificities for ssDNA sequences. We then demonstrate potential applications including ssDNA detection and digestion of ssDNA from Rolling Circle Amplification (RCA). Because of their diverse set of ssDNA cleavage specificities, we anticipate that the widely distributed Ssn nuclease family may be harnessed as molecular tools to selectively modify single-stranded DNA in multiple biotechnological applications.

Results

Characterization of NTS repeats in *Neisseria* and other bacteria
The dRS3 element was originally described as the 20 bp palindromic repeat ATTCCC[N$_8$]GGGAAT, present in hundreds of copies in the pathogenic *Neisseria* genomes[28,32]. We observed that this perfectly conserved repeat masked, in nearly 50% of dRS3 copies, another more complex sequence motif of around 30 bp (Fig. 1a). Defined as CGTCATTCCCGCGMAVGCGGGAATCYRG, we term this element *Neisseria* Transformation Sequence (NTS). While it shares the same palindromic motif as the dRS3 (underlined), the NTS distinguishes itself

from the former with a characteristic orientation of flanking non-palindromic sequences. The NTS motif is widespread in *Neisseria* spp. genomes (Supplementary Data 1), but the number of repeats varies significantly between species (Fig. 1b). A striking overrepresentation of the NTS is observed in the phylogenetic clade encompassing the only two human pathogens, namely *N. meningitidis* (*Nm*) and *N. gonorrhoeae* (*Ng*). We then mapped the NTS repeats within the *Nm* genome, revealing a non-random distribution with clusters around genes, including some encoding for well-known virulence determinants (Fig. 1c, Supplementary Data 2). A similar distribution is observed in *Ng* (Fig. S1a). To ascertain whether this was a *Neisseria*-specific phenomenon, we first searched for NTS repeats in genomes from other bacterial families using BLAST[33]. We uncovered many species harboring NTS-like repeats such as *Rhizorhabdus wittichii* and *Wolbachia pipientis*, suggesting a more widespread phenomenon (Fig. 1d, e).

SRM motifs are associated with a specific gene (*ssnA*) in multiple bacterial genomes
To explore the function of these motifs, we first hypothesized that bacterial proteins, such as regulators, might bind to or interact with the NTS repeats to fulfill specific biological role(s) in several species. Additionally, we hypothesized that if proteins do interact with the NTS, the genes encoding these proteins should be found 1) mostly in bacterial genomes harboring NTS and 2) be spatially associated with NTS elements (for example to be possibly transcriptionally regulated by them). Using in-house scripts[34], we searched for NTS motifs in the flanking regions of genes encoding the closest homologs of all *Nm* proteins (NMV_0001 to NMV_2374) in each bacterial genome from the representative/reference RefSeq database (excluding *Neisseriaceae*) (Fig. 2a–c, Supplementary Data 3). We detected a specific set of genes, homologous to *NMV_0044* in *Nm* (and renamed *ssnA* herein) that was found to have 1 NTS (or more) repeat in its flanking regions in 31% of homologs analyzed and flanked by at least 1 NTS repeat in both the 5' and 3' regions in 14% of homologs (Fig. 2a, Supplementary Data 3). Inversely, a homolog of the *ssnA* gene is commonly found in species harboring numerous copies of the NTS repeats, across great phylogenetic diversity of bacteria (Supplementary Data 4).

Despite their presence in diverse bacteria, some *Neisseria* genomes that contained a homolog of *ssnA* did not have NTS elements. We therefore conducted a de novo search for motifs in the genomic regions flanking the *ssnA* homolog in *Neisseria wadsworthii* and related species and identified a similar but distinct hairpin-forming repeat, differing only by a few well-conserved substitutions, which we term NTS$_{var}$ (Fig. 1a, b, Fig. S1c). This prompted us to build a methodology to further capture diversity in NTS-like elements. We examined genomic regions flanking the regions encoding homologs of SsnA from 29 diverse bacterial genomes from 4 phyla (Supplementary Data 5, Fig. 2c). We then aligned NTS-like elements to refine a degenerated search sequence for NTS-like elements, which we term Ssn-related motifs (SRM). We repeated the searches with this degenerated motif and obtained an even more convincing association: 71% of regions surrounding genes encoding SsnA homologs harbor one or more SRM sequences and 42% of homologs are flanked by at least 1 SRM repeat in both the 5' and 3' regions (Fig. 2a, Supplementary Data 3). Finally, 81% of genomes with >200 SRM repeats (137/169) have a region encoding a homolog of SsnA (by tblastn criteria E-value < 1e-10, Supplementary Data 4). This clearly demonstrates the strong association of Ssn encoding genes with their SRM motif in multiple genomes from a diversity of phyla.

SsnA from *N. meningitidis* is a specific single-stranded DNA nuclease that cleaves near the NTS
The *ssnA* gene from *Nm* encodes a single-domain hypothetical endonuclease of 94 amino acids belonging to the uncharacterized cd10448 family in the GIY-YIG protein superfamily. SsnA proteins share the
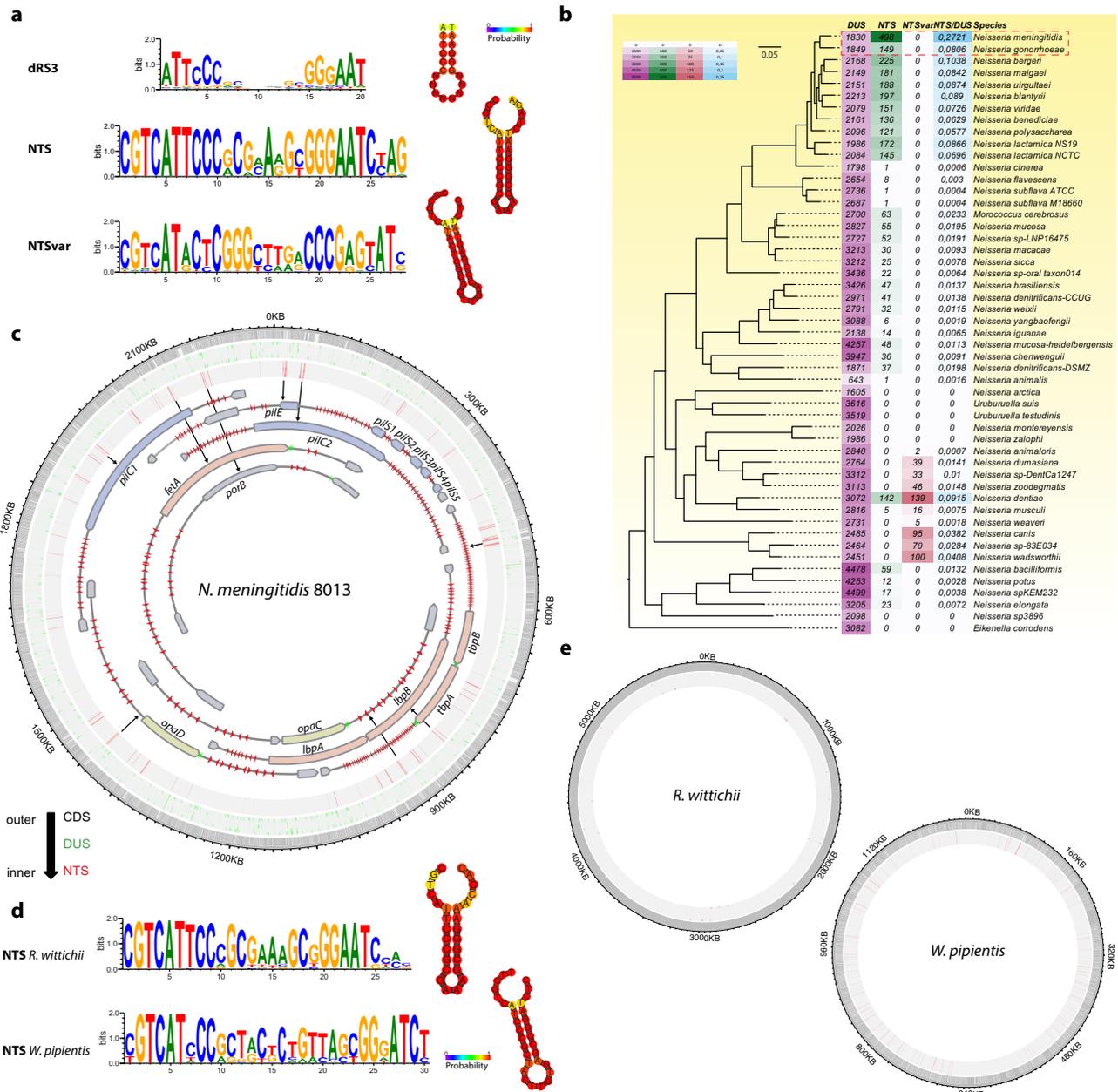
Fig. 1 | Hairpin repeats surround key genes in *Neisseria* and other bacteria. **a** Logos and predicted secondary structures of repeated sequences from *N. meningitidis* (dRS3, NTS) and *N. wadsworthii* (NTS$_{var}$). For the secondary structure, the different colors represent the base-pair probability. **b** Core-genome phylogeny of *Neisseria* species with their associated repeated elements. The four color gradients represent the relative abundance of the different enumerated elements. **c** Circular map of the *N. meningitidis* 8013 genome with its DUS (green) and NTS (red) repeats depicted. Major clusters of NTS are highlighted with black arrows pointing to the corresponding genomic landscape. Those major clusters are emphasized with schematic representations of the loci of interest (not drawn to the scale). **d** Logos and predicted structures and (**e**) whole-genome distribution of SRM repeats in *R. wittichi* and *W. pipientis*.

characteristic 3D fold of the GIY-YIG domain (Fig. S1d, e, Supplementary data 10). Unlike SsnA, the other characterized members of the GIY-YIG superfamily contain additional functional domains or extensions which give these proteins their functional specificities (reviewed in ref. 3).

In order to characterize the enzymatic activity of SsnA, we expressed and purified the *Nm* protein in *Escherichia coli* along with variants mutated in the predicted active sites (Fig. S2a, b). EMSA (gel-shift) assays were performed to detect DNA-binding activity of the protein, while nuclease assays were used to detect DNA-cutting activity, both using fluorescently labeled DNA (Fig. 3). No enzymatic activity

was detected with a dsDNA substrate, including genomic DNA, plasmid DNA and PCR products. Similarly, no activity was observed when SsnA was incubated with dsDNA harboring the NTS repeat (Fig. 3a). However, a specific binding and cutting activity was obtained when the same sequence was tested as single-stranded DNA (Figs. 3a and S2c) liberating a 20 nucleotide fragment (Fig. S2d) allowing us to map the cleavage site. Importantly, ssDNA devoid of NTS or with the reverse strand of the NTS was not cleaved (Fig. 3a). A GST-tagged SsnA showed a similar activity, confirming that the enzymatic activity is likely not affected by the purification tag (Fig. 3a). Moreover, directed mutagenesis of the conserved tyrosines from the GIY-YIG motif as well as a
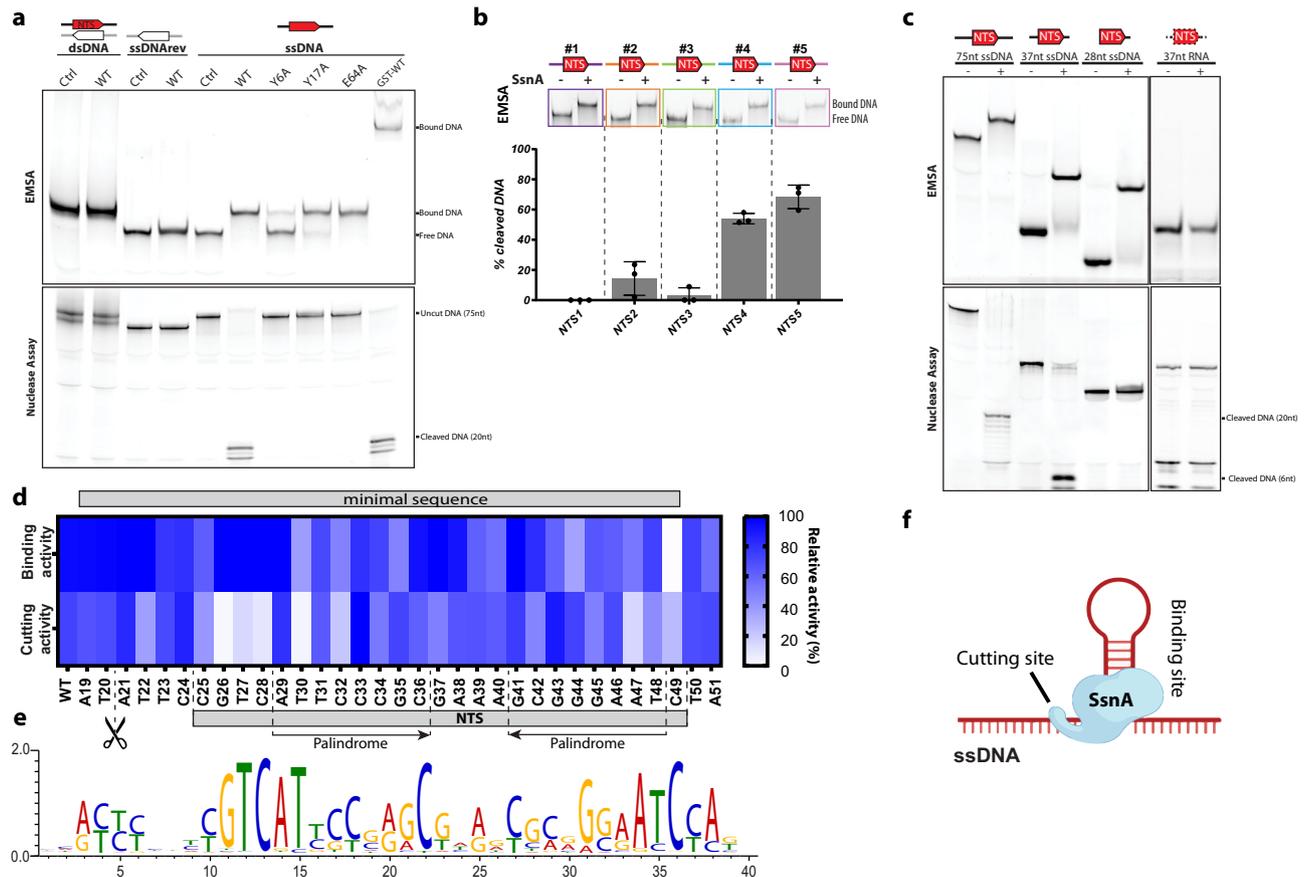
Fig. 2 | NTS repeats are associated with hypothetical GIY-YIG nucleases.
a Distribution of homologous-protein encoding genes of the *Nm* proteome and the fraction of such homologs with NTS or SRM repeats flanking their genes at both sides. Each point represents one protein in the *Nm* proteome, and the plot relates the number of homologs found in the NCBI reference / representative genome database to the fraction of those homologs flanked by SRMs. **b** Distribution of the NTS repeats (red) near selected *ssnA* homologs (blue). **c** Sequence alignment of *ssnA*-associated NTS repeats from selected species, in the same order as in (**b**). Darker blue indicates better conserved nucleotides.

---

predicted cofactor binding site abolished any nuclease activity while permitting binding of ssDNA (Fig. 3a). Because the NTS repeats are not perfectly conserved, we then assessed whether SsnA could interact with different copies from the *Nm* genome. Five different copies were tested as ssDNA, each with their respective flanking sequences (Fig. 3b). Although all of them were efficiently bound by the protein, they were not all cut with the same efficiency, suggesting complex and distinct binding and cutting specificities. To better understand this enzyme's specificity, we proceeded to test truncated versions of an NTS-containing oligo that was efficiently cut (Fig. 3c). Notably, cutting activity was observed with a sequence of 37nt (6nt before the cleavage site and the NTS) or longer, while the 28nt sequence (3nt after the cleavage site and the NTS) only allows for binding of SsnA. We performed cleavage assays with oligos starting from 2nt before the cleavage site (our nuclease assay's limit of detection) and varying lengths of 3' overhangs (Fig. S2e). This allowed us to identify the minimal sequence for which cleavage can still be observed (as schematized Fig. 3d). Finally, we have verified the activity on RNA and, as with dsDNA, SsnA is unable to interact with RNA, confirming its classification as a specific ssDNase (Figs. 3c and S2f).

To assess the importance of each base in the NTS sequence within NTS75, we conducted a series of EMSA and nuclease assays on sequences with single-point mutations. These assays were quantified and summarized in Fig. 3d, with detailed quantification provided in Fig.

S2g, h. Mutations in the palindromic region of the NTS resulted in lower binding activity, but the most crucial nucleotide for binding was shown to be the cytosine immediately after the palindrome. Interestingly, cleavage activity was primarily affected by the three nucleotides upstream of the palindrome, confirming distinct specificities for both enzymatic activities and the importance of the NTS orientation. Mapping the precise cleavage site of SsnA revealed that this region can support some variation and still be cut efficiently (Figs. 3d and S2g, h). The analysis of residue conservation in the NTS across evolution in all organisms carrying an Ssn homolog (Fig. 3e) mirrors the results obtained for the directed mutagenesis experiments of the NTS in *Nm* (Fig. 3d). Furthermore, we observed that modifying the palindromic sequence while preserving the hairpin structure results in a loss of SsnA activity (Fig. S2i, Supplementary data 9). This observation suggests that recognition is not solely mediated by the structure but also by the sequence.

In light of these observations, we conclude that SsnA from *N. meningitidis* is a site-specific single-stranded endonuclease capable of binding to the NTS as ssDNA only and cutting a few nucleotides upstream in the non-conserved variable flanking region (Fig. 3f). This study being the initial characterization of SsnA, we proceeded with a more thorough characterization of the enzyme's catalytic activity, revealing its optimal temperature, cofactor requirements, and kinetic parameters (Fig. S3). Our results indicated that magnesium is the most

Fig. 3 | SsnA is a specific single-stranded endonuclease targeting NTS repeats. **a** SsnA binding (top) and cleavage (bottom) activities on fluorescently labeled oligos containing an NTS repeat (oligos NTS75_dsDNA, NTS75rev, and NTS75). His-tagged WT and mutated SsnA (Y6A, Y17A, E64A) were assayed along with a GTS-tagged WT version. Each condition of the representative experiment was performed at least three times. **b** SsnA binding (top) and cleavage (bottom) activities on 75 nt ssDNA oligos containing NTS repeats from different regions of the *Nm* 8013 genome (oligos NTS1-5). The cleavage product was quantified from three independent experiments (mean ± SEM). **c** As in a, but with truncated versions of the ssDNA substrate (oligos NTS75, NTS37, NTS28), along with a 37 nt single-stranded RNA of the same sequence (oligo NTS37_RNA). Each condition of the representative experiment was performed at least three times. **d** Effect of individual mutations along and around the NTS repeat on the ssDNA-binding (top) and cleavage (bottom) activities of SsnA. Binding and cleavage products were quantified from at least three independent experiments. The cut site and NTS location are indicated. The blue gradient represents the relative activity. **e** The logo represents residue conservation within the NTS, identified through de novo analysis of regions flanking homologs of SsnA performed in the NCBI representative/reference database. **f** Illustration of the predicted interaction of SsnA with NTS-containing ssDNA. Image created in BioRender. Veyrier, F. (2025) https://BioRender.com/x76q784. Source data are provided as a Source Data file[76].

effective metal cofactor, with an optimal activity temperature of 37 °C, and that the SsnA/NTS75 complex has a Kd of 4.5×10^{-9} M under our experimental conditions (Fig. S3).

SsnA from *N. meningitidis* modulates natural transformation

Only a handful of cellular processes involve single-stranded DNA, such as replication, recombination, and repair. In *Neisseriaceae* and other naturally competent bacteria, ssDNA is also of importance during transformation, one of the major modes of horizontal gene transfer (HGT). Although exogenous DNA is double-stranded in the environment, the DNA uptake machinery of competent species translocates only one strand to the cytoplasm[35]. To assess whether the specific ssDNase activity of SsnA contributes to this process, we generated *ssnA* knockout (KO) strains of *Nm* 8013 and *ssnA* complementation strains (Compl) ectopically expressing *ssnA* under a strong promoter, resulting in overexpression. Proper deletion, complementation and normal fitness of the bacteria were verified (Fig. S4).

Since SsnA specifically cuts ssDNA near NTS repeats in vitro, we assessed whether the presence of SsnA would impact frequency of chromosomal integration of transforming DNA with NTS in *Nm*. When transforming *Nm* strains with a plasmid harboring several NTS repeats

in the recombinant regions, the *ssnA* deletion strain exhibited a 5-fold increase in transformation rate, while the complementation strain was transformed at an even lower rate than the WT strain (Fig. 4a). Similar results were obtained when transforming *Nm* with additional NTS-containing plasmids that integrate elsewhere in the genome, indicating that this effect is not restricted to one locus (Fig. S5a–c). We also assessed whether the SsnA deletion strains had the same overall competence levels when using an integrative plasmid free of NTS (Figs. 4b and S5d–f). Notably, in this setting, the *ssnA* deletion strain showed no major differences with the WT strain in transformation rate. To remove any potential sequence bias and prove that the observed effect of SsnA on natural competence is indeed due to the presence of NTS repeats in the transforming DNA, we designed plasmids and recipient *Nm* strains to have identical sequences throughout the recombination regions but with two mutations (in all NTS repeats) shown to prevent binding and cutting in vitro (G26 and C49 as seen in Fig. 3d). When transforming *Nm* with the intact NTS-containing plasmid, the loss of SsnA was again accompanied by an increase in transformation rate (Fig. 4c). However, this phenotype was lost when the same assays were performed with the mutated NTS-containing plasmid, confirming these repeats' role in the integration of exogenous
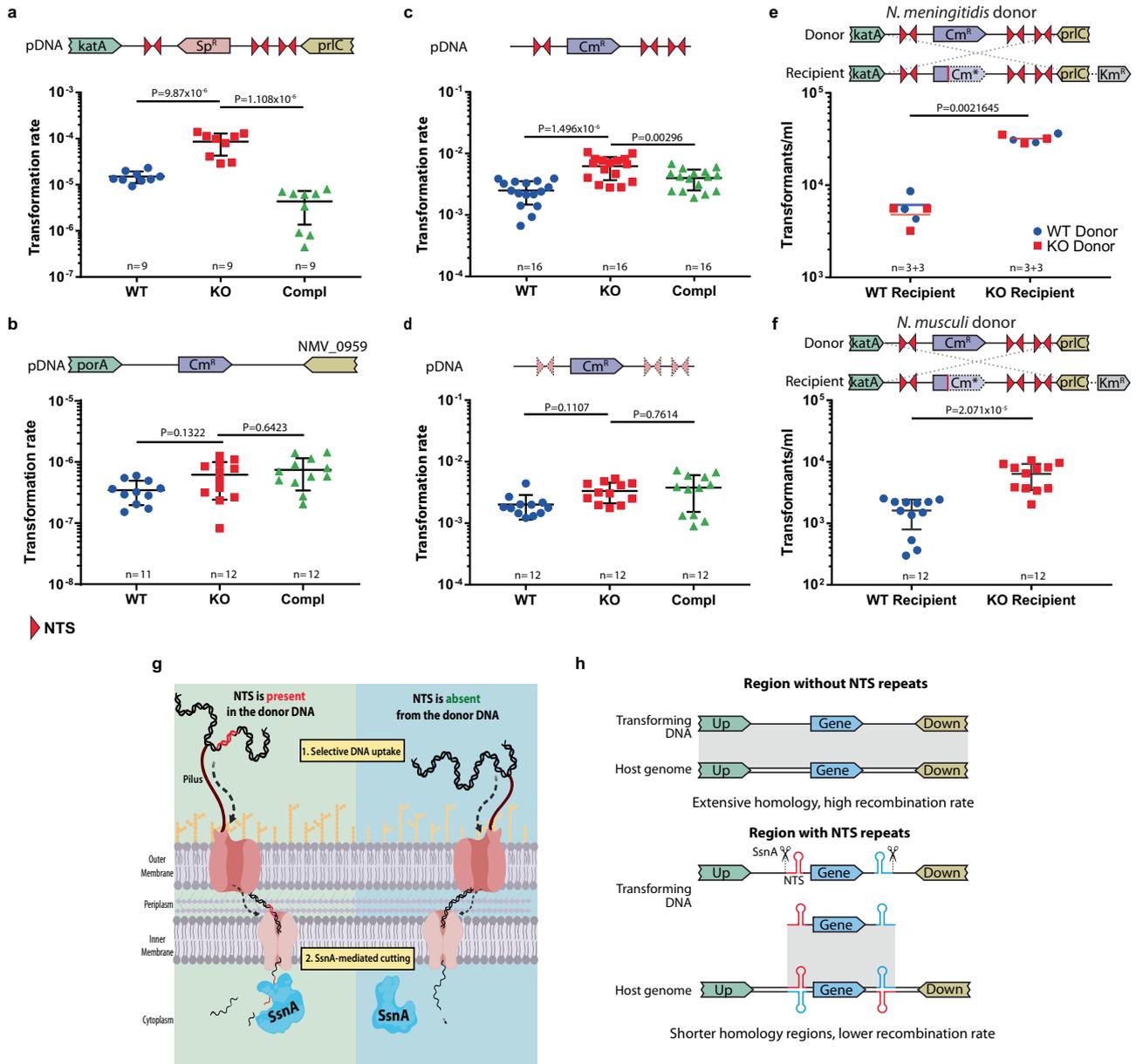
Fig. 4 | SsnA modulates natural transformation in *N. meningitidis*. Transformation rates of *N. meningitidis* strains using a plasmid (**a**) with NTS repeats (pKOYebN::Sp), (**b**) without NTS repeats (pCompInd), (**c**) with NTS repeats and limited flanking regions (p5'3'inCm), (**d**) with the same DNA as (**c**) but with mutated NTS repeats (p5'3'inCmMut2). **e** DNA exchange between *N. meningitidis* co-culture with another *N. meningitidis* donor strain. **f** As in e, but where the donor is the commensal species *N. musculi*. Mean of at least three biological replicates, ± SD, is reported for all transformation assays. One-way anova with post-hoc tukey's multiple comparisons test was performed on transformation tests. Co-cultures were analyzed using two-tailed Mann-Whitney tests. **g** Illustration of the proposed interplay between SsnA and the NTS repeats during natural competence of *N. meningitidis*. Each NTS is illustrated with a red arrow, which also indicates its orientation. Transforming DNA enters the cytoplasm as single-stranded (ssDNA). The SsnA nuclease then cuts this ssDNA at specific sites near NTS repeats, modulating its integration to the host genome. **h** Proposed mechanism for the SsnA-mediated modulation of DNA integration to the host genome. When the transforming ssDNA recombines to the host genome in a region devoid of NTS repeats, the extensive homology allows it to be integrated at a high frequency. However, if the transforming ssDNA recombines to a region harboring NTS repeats, SsnA will specifically cleave it near the repeats, shortening the homology region and hence reducing the rate of homologous recombination. Image Created in BioRender. Veyrier, F. (2025) https://BioRender.com/x73y116. Source data are provided as a Source Data file[76]. Cm^R: Chloramphenicol resistance gene.

DNA (Fig. 4d). Using whole genomic DNA with the same features produced identical results, suggesting that this effect is independent of the length of the homologous regions or the methylation state of the donor DNA (Fig. S5c).

To better approximate the true biological conditions of horizontal gene transfers and the role of NTS and SsnA, we performed a similar set of experiments using co-cultures of donor and recipient strains instead of purified DNA. Similarly to transformation assays with purified DNA, the SsnA knockout recipient *Nm* acquired the selection marker much more efficiently than the WT recipient *Nm* (Fig. 4e). Whether or not the donor *Nm* expressed *ssnA* had no influence on this phenotype, indicating that SsnA acts solely on the integration of transforming DNA and not on its donation. To assess whether this mechanism is also active in cross-species HGT, we repeated the assay using the commensal *Neisseria musculi* as the donor, which yielded similar results (Fig. 4f).

Together, these results indicate that the specific ssDNase activity of SsnA is used in vivo by *Nm* as an ssDNA restriction mechanism. In the example tested here, as illustrated in Fig. 4g, cleavage of incoming ssDNA based on the presence and localization of NTS repeats may result in fragments with shorter homologous regions to the host genome, thus affecting their integration by homologous recombination (Fig. 4h).

SsnA belongs to an extensive family of sequence-specific ssDNases (Ssn)

After doing a comprehensive characterization of the SsnA protein in *Nm*, we wondered if the *ssnA* gene had any homologs that could share a similar enzymatic activity. SsnA is classified in the conserved-domain database[36] (CDD) as belonging to the cd10448 family. To determine if proteins sharing its characteristics belonged to a monophyletic family, we conducted a phylogenetic analysis of all GIY-YIG domain-containing proteins ($n = 46,332$, clustered by sequence identity into 8528 clusters) in our database of reference/representative bacterial genomes (Fig. 5a). Plotting homology to the *Nm* SsnA protein, protein length, presence of SRM repeats in flanking genomic regions, and CDD domain assignments, we identified a clade of representative proteins in the larger GIY-YIG superfamily phylogenetic tree that is nearly synonymous with the distribution of the cd10448 (GIY-YIG unchar_3) domain. To our knowledge, none of those proteins are comprehensively characterized and most of them are considered hypothetical.

Complementing this analysis to focus on the taxonomic distribution of Ssn proteins, we used the full NCBI CDD database to generate an extensive list of thousands of presumed homologous single-domain cd10448-containing GIY-YIG proteins of similar size. All sequences were aligned and used to infer a maximum-likelihood phylogeny (Fig. 5b). While most of the SsnA homologs are found in gram-negative bacteria (mainly Pseudomonadati), they are widely distributed throughout across bacterial phyla.

To demonstrate the ssDNA-specific activity and possible differences in specificity, we purified several homologs of SsnA from *N. elongata*, *N. wadsworthii* and *W. pipientis*. Since each protein is flanked by its own NTS repeats with various polymorphisms and different flanking regions, each homolog was assayed with its associated sequence. As with the *Nm* SsnA, each homolog was confirmed to specifically cleave near its respective SRM repeat as ssDNA (Fig. 5c–f). Since some of the tested sequences were similar, we then assayed each homolog with every SRM sequence (Fig. 5g–j). Although some cross-specificity is observed, the sequence from *W. pipientis* is only cut by its own SsnA homolog (Fig. 5j), which is the most distant as seen in Fig. 5a. The SsnA homologs behave differently in the presence of the same sequence. For example, the $NTS_{N.elongata}$ is cut one nucleotide before by $SsnA_{N.elongata}$ as compared to the other SsnA (Fig. 5h). The $NTS_{var (N.wadsworthii)}$ is cut with different efficiency and at different positions between the homologs (Fig. 5i, *Nw* and *Wp*). Together, these results confirm that members of the widespread Ssn family (cd10448: GIY-YIG_unchar_3 protein family) are specific single-stranded endonucleases hence supporting the identification of SsnA proteins as a distinct functional enzyme family with a diverse set of cleavage specificities.

Proofs of concept for potential applications in biotechnology

Although this work is a preliminary characterization of Ssn homologs in four bacteria (*Nm*, *N. elongata*, *N. wadsworthii* and *W. pipientis*), this site-specific ssDNase activity opens the door to numerous applications in the field of ssDNA molecular biology[37].

(1) Concatemeric ssDNA digestion. In order to demonstrate potential uses, we first performed a Rolling Circle Amplification (RCA) of a circularized 120 nt ssDNA molecule (NTS75T7F) containing the reverse NTS. This generated a continuous concatemeric ssDNA molecule containing the forward NTS (Fig. 6a left panel). Following a digestion by SsnA and staining of the gel using ssDNA-specific dye (ssGreen-Lumiprobe), we observed that this generated the expected 120 nt digestion product (Fig. 6a right panel). As a negative control, we did not observe a digestion product for the RCA of the circularized reverse sequence (NTS75T7R). This molecule, which contained the forward sense NTS, generated a continuous concatemeric ssDNA molecule containing the reverse NTS without cleavage. This emphasizes the straightforward use of an Ssn endonuclease for site-specific cleavage of an RCA product without intermediate steps such as those dependent on dsDNA or nickase enzymes.

(2) ssDNA-programmable digestion. Beyond the obvious utility of simple cleavage of ssDNA and because we observed that nucleotide identity in the loop of the hairpin was not essential for cleavage, we hypothesized that we could remove the loop from the hairpin stem loop while retaining enzymatic activity. We therefore tested the activity of the enzyme in the context of two ssDNA oligonucleotides: each containing one-half of the NTS that form the stem of the hairpin without the loop segment (Fig. 6b left panel) and demonstrate cleavage (Fig. 6b right panel). This result constitutes a proof of concept for inducible/programmable cleavage of ssDNA. Revealing SsnA/NTS modularity, we also show that we can add different ssDNA extensions to the loop position (Fig. 6b). This extension can be either non-complementary (such as the first half oligo HemiF, the second half oligo $39\text{-}77_{polyA}$ or both as exemplified in the 3rd, 4th and 5th schemas Fig. 6b left panel) or two complementary random extensions which increase the length of the hybridized stem (oligo LT, 6th schemas Fig. 6b left panel) of the former hairpin.

We next performed an RCA on a circularized oligonucleotide (HemiR) which generated a continuous concatemeric ssDNA molecule containing only the first half of the NTS (Fig. 6c left panel). To this RCA, we subsequently hybridized an oligonucleotide containing the second half of the NTS before performing a nuclease assay. This resulted in the digestion of the RCA product into a single fragment of the size of the of the original HemiR oligonucleotide (Fig. 6c right panel). This experiment demonstrates that the NTS/Ssn system could be used for programmable cleavage using a DNA-guide analogous to Prokaryotic Argonaute proteins[38,39] or CRISPR/cas systems[40].

(3) Detection/Quantification of ssDNA. Building on our previous findings on programmable cleavage, we added another layer of applicability using a fluorescent quencher system that could be used for ssDNA quantification/detection. In Fig. 6b and Fig. 6c, we have shown that cleavage can be triggered after hybridization of two different DNA fragments containing respectively the two halves of the NTS (which may reform the stem). This finding led us to develop a detection probe containing the first half of the NTS (cleavage site and half of the palindrome) along with a 5′ fluorophore and a 3′ quencher (schematized in left part of Fig. 6d). This probe can therefore be used to detect the presence of a ssDNA sequence that carries the other half of the NTS. Using the two ssDNA oligos consisting of the split NTS described above, we were able to measure the release of the fluorophore from the quencher during ssDNA cleavage by SsnA. When the probe hybridizes with the complementary palindrome, SsnA cleaves the probe, thereby releasing the fluorophore from the quencher, and allowing for the quantification of the target DNA (herein $39\text{-}75_{polyA}$ in Fig. 6d). This method offers a highly sensitive means of detecting specific ssDNA. However, we further optimized this detection strategy by coupling it with RCA. In this case, we circularized an oligo containing the complementary sequence of the second half of the NTS and performed isothermal amplification of the target sequence (second half of the NTS as schematized in red in the left part of Fig. 6e). This
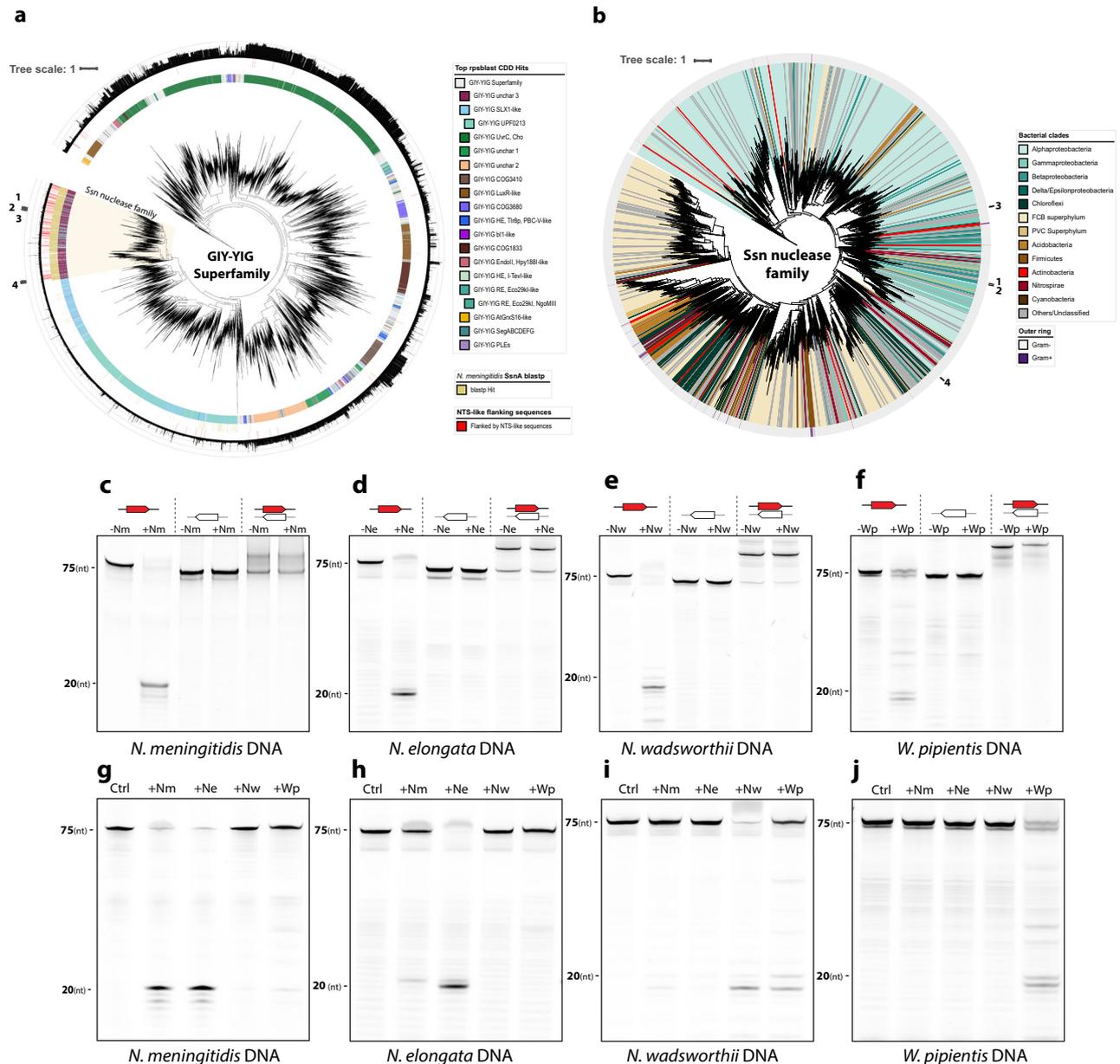
Fig. 5 | SsnA belongs to a broad family of ssDNases with diverse specificities.
a Maximum-likelihood phylogenetic analysis of all GIY-YIG domain-containing proteins in the bacterial reference/representative database. From innermost to outermost, annotation rings indicate: the top GIY-YIG domain hit in the custom conserved domain database (CDD) determined by rpsblast, the distribution of blastp hits among these proteins with the *N. meningitidis* SsnA protein as a query (*E*-value $< 1 \times 10^{-10}$), the proteins for which SRM elements are found in the genomic region flanking the coding sequence of the corresponding gene, and finally, protein lengths, with scale lines at 250, 500 and 750 amino acids. The proposed Ssn nuclease family clade is shaded. Numbers indicate the representative SsnA

homologs closest to 1- *N. meningitidis* (*Nm*) 2- *N. elongata* (*Ne*) 3- *N. wadsworthii* (*Nw*) 4- *W. pipientis* (*Wp*). **b** Maximum-likelihood phylogeny of the Ssn nuclease family. Inner ring colors indicate the host genome's classification into major bacterial clades, while the outer ring indicates the Gram stain classification. c to f, nuclease activity of purified SsnA from (**c**) *Nm*, (**d**) *Ne*, (**e**) *Nw*, (**f**) *Wp*, on an NTS-containing ssDNA from their respective genomes. g to j, nuclease activity of purified SsnA from *Nm*, *Ne*, *Nw* and *Wp* on ssDNA from (**g**) *Nm*, (**h**) *Ne*, (**i**) *Nw*, (**j**) *Wp*. Source data are provided as a Source Data file[76]. All experiments were performed at least three times (or more).

Discussion

We first report here that SsnA is a specific single-stranded DNase that modulates integration of transforming DNA (containing the NTS) in *Neisseria meningitidis*. The dRS3 repeats, found in much greater numbers in pathogenic *Neisseria* species compared to their

commensal counterparts, have generated many hypotheses since their description more than two decades ago. In silico analyses of their distribution showed that orthologous coding sequences (CDS) from several *N. meningitidis* genomes adjacent to dRS3s have an average percentage identity lower than the average percentage identity of orthologues not in proximity to dRS3s. It was suggested that dRS3-adjacent genes have increased sequence diversity explained by the fact that these repeats could be used as homologous region for recombination[31,41,42]. Revisiting these repeats in a large set of

Fig. 6 | Example of potential SsnA/NTS biotechnological applications with or without coupling rolling circle amplification (RCA). In this figure, the left panels in each sub-figure illustrate the experimental setup, while the right panels illustrate the results. **a** SsnA cleavage of an NTS75 concatemer amplified by RCA at different time points. Semi-digested concatemers of various sizes, containing multiple NTS75 sequences, are incrementally digested, ultimately yielding a single band matching the original oligonucleotide size (120 nt) but corresponding to its reverse sequence. **b** SsnA activity on two ssDNA oligonucleotides, each containing half of the NTS stem without the loop. After hybridization with sequences containing the complementary half of the palindrome, cleavage is observed via post-electrophoresis staining (ssGreen-Lumiprobe). The "1–38" sequence includes the first 38 nucleotides of NTS75, while "39–75" includes the second half of NTS75. "39–75polyA" represents 39–75 with added polyA (in red), "HemiF" consists of the 1–38 sequence with random ends (yellow), and "LT" contains the 39-75 part of NTS with a complementary HemiF sequence to increase stem size. An asterisk (*) indicates 5′ 6-FAM labeling, and "HI" corresponds to heat-inactivated samples.

c Cleavage of an RCA-produced ssDNA concatemer containing the first half of the NTS. HemiR circularization and amplification produce HemiF. Hybridization with the 39–75 oligo enables specific cleavage using an ssDNA guide. **d** Quantification assay via fluorescence quenching. An ssDNA sequence (*1–38Q) containing the first half of the NTS with 5′ 6-FAM and a 3′ quencher (3′ BHQ1) was hybridized with the 39–75 sequence and cleaved by SsnA to release fluorescence, quantifiable based on the amount of 39–75 oligo. **e** Detection of the 39–75 region via RCA amplification, hybridization with *1–38Q, and cleavage by SsnA over 150 min. The RCA with a correctly oriented 39–75 concatemer is detected, while the reverse concatemer is not. The indicated concentrations represent the amount of DNA originally used for RCA. Histograms represent the mean, and orange bars represent the standard deviation. Each point represents an independent reaction mix. Three replicates were done with DNA and protein, two without protein. Experiments were repeated with SsnA from two purification batches. Schematic representation Created in BioRender. Veyrier, F. (2025) https://BioRender.com/x76q784 and https://BioRender.com/y23w368. Source data are provided as a Source Data file[76].

commensal and pathogenic *Neisseria* genomes, we demonstrate a subset of these repeats are in fact longer NTS repeats. In addition to the semi-palindrome forming the hairpin core of these repeats, NTSs share a highly conserved upstream sequence that gives them an

orientation and is crucial to their interaction with SsnA. Our transformation assays demonstrate that the specific nuclease activity of SsnA acts as a single-stranded DNA restriction mechanism which could influence the rate of integration of transforming DNA depending on

the presence of NTS repeats. Taken together, these data demonstrate that NTS repeats are part of a complex regulatory mechanism for HGT occurring through natural transformation, in addition to the already described DUS system. Herein, transforming DNA with NTS-Gene-NTS configuration has the opposite effect on transformation rates compared to the DNA Uptake Sequences (DUS), which selectively promote DNA import through direct interaction with the host pilus[43,44]. In addition, their distribution does not seem to be the same nor to be random. NTS are rarely inside genes (2% in *N. meningitidis* 8013, GCA_000026965.1) unlike DUS that are often found within CDS (34%). DUS repeats are enriched in genes for genome maintenance (DNA replication, recombination, repair)[45], whereas NTS form important clusters frequently adjacent to genes encoding membrane proteins often involved in host interaction (Fig. 1c; Supplementary Data 2). To explain the effects on transformation, we emitted a first hypothesis that this enzyme may regulate uncontrolled and potentially detrimental recombination with exogenous DNA that could inactivate, delete or replace those important but often non-essential genes. Supporting this hypothesis is the observation that the presence of commensal *Neisseria* induces death of *Ng* when their homologous DNA is acquired by *Ng* using competence[46]. This suggests that in certain contexts, exogenous DNA integration can be detrimental. On the other hand, *Neisseria meningitidis*, which is deficient for some DNA repair systems (SOS response, photoreactivation system, response to alkylative damage[47]), may employ its natural competence to repair its DNA via homologous recombination[48,49]. A second hypothesis concerns genes encoding membrane proteins surrounded by NTS elements, (Fig. 1c; Supplementary Data 2) that are often antigenic and under the selective pressure of hosts' immunity. These genes may be less effectively repaired, leading to increased sequence variability and thus antigenic drift. Finally, it is also possible that alternative configurations of NTS elements relative to the position of gene(s) in the transforming DNA could have other modulating effects on recombination.

The biological role(s) of Ssn ssDNA nucleases in other bacteria remain to be elucidated. Interestingly, in the species that have more than 300 SRM elements, there is a clear enrichment of plant and animal symbionts (such as *Rickettsia, Wolbachia, Neisseria, Sphingomonas, Povalibacter* and *Devosia*). The intriguing gene architecture NTS-*ssnA*-NTS is identical to mobile genetic elements such as insertion sequences (IS). It has been shown that an atypical family of insertion sequences, the IS200/IS605, encodes transposases with dsDNA nicking activity[50] that specifically target ssDNA hairpins[51]. Just like SsnA, these transposases are closely associated with repetitive extragenic palindromes (REP) that cluster into larger structures called bacterial interspaced mosaic elements (BIMEs). It is thought that such transposases were "domesticated" by their hosts to control other functions such as generating genomic diversity, in addition to being responsible for the accumulation of their respective DNA repeats[52,53]. These intriguing similarities with SsnA and the NTS repeats raise the hypothesis that this genetic element (NTS-*ssnA*-NTS) was domesticated by *Neisseria* or other symbiotic species to perform important biological functions linked with genome dynamics (such as controlling HGT). This scenario may have led to the massive expansion of NTS repeats in some species.

We secondly report here the in vitro activity of Ssn proteins as site-specific ssDNA endonuclease. SsnA belongs to a broad family of single-domain hypothetical proteins belonging to the GIY-YIG superfamily with the majority of the genes encoding these homologs are flanked by SRM elements (20−50 nt inverted repeats likely forming a hairpin secondary structure). This suggests a vast number of NTS-Ssn pairs, with their own co-evolutionary dynamics that will take time to characterize in detail. This complex interaction is remarkable for such a small, monomeric, single-domain protein. In contrast, nearly all characterized GIY-YIG proteins are multi-domain proteins with a defined DNA-binding domain in addition to the GIY-YIG nuclease

domain[42]. These important differences justify the classification of Ssn nucleases as a novel family in the GIY-YIG superfamily, now comprising Ssn (cd10448), MSH1, Slx1, homing endonucleases, Penelope elements, UvrC, COG3680 and COG1833[54,55]. Structural studies will be of great interest to investigate SsnA's intricate interaction with ssDNA and further explore its potential applications in biotechnology. Nevertheless, ssDNA-specific endonucleases are quite uncommon (reviewed in ref. 56). The few enzymes capable of cutting single-stranded nucleic acids also act on dsDNA or RNA or show no sequence specificity. SsnA is unique in the sense that no binding nor cutting could be detected with both dsDNA and RNA. In addition, the NTS is directional as its ssDNA reverse strand is not cut by SsnA. We have demonstrated in vitro that this enzyme possesses a complex binding and cutting specificity, where it binds to NTS repeats to cleave ssDNA upstream. Another advantage of the Ssn/NTS system is that cleavage can also happen in the presence of two ssDNA oligos that harbor each half the NTS sequence. Although this programable cleavage mechanism is comparable to other commercially available nucleic acid-guided nuclease systems (such as Prokaryotic Argonaut or CRISPR/cas), the Ssn/NTS system offers the advantage of a vast array of potential paired sequence/enzyme specificities that could be used combinatorically for direct ssDNA cleavage or DNA-guided cleavage. Additionally, the coupling of RCA technique with the specific cleavage of SsnA could enable the development of particularly sensitive ssDNA detection methods. For all these reasons, we anticipate that these unique and flexible properties will form the basis of a suite of molecular tools that may offer alternatives to existing technologies or open the door to future ssDNA-based applications.

In conclusion, we report important insights on the evolution of the pathogenic *Neisseria* species. Furthermore, the discovery and characterization of the SsnA protein and its family of endonucleases provide an important class of tools for molecular biology and biotechnology research. ssDNA has numerous applications in basic research, molecular diagnostic, and genetic engineering. Once the sequence specificities of a broader set of Ssn nucleases are characterized, this unique enzyme family may serve as versatile molecular tools to manipulate ssDNA.

Methods

Neisseriaceae motif analysis

A set of 75 *Neisseriaceae* genome assemblies was collated from the NCBI and from our own database (Supplementary Data 1). The genome database was then queried for ATTCCCNNNNNNNNGGGAAT (dRS3), CGTCATTCCCGCGMAVGCGGGAATCYRG (NTS), CGTCA-TACTCGGGNNNNNCCCGAGTATC (NTS$_{var}$) using fuzznuc from the EMBOSS suite (v6.6.0.0)[57]. NTS and NTS$_{var}$ were queried with one mismatch permitted and dRS3 with no mismatches permitted.

Sequence logos were generated from the fuzznuc hits of defined genomes using WebLogo 3 after alignment with MAFFT (v7.3.0)[58], allowing three mismatches from the query[59]. The dRS3 and NTS logos derive from *N. meninigitdis* 8013, while the NTS$_{var}$ logo derives from *N. wadsworthii* DSM 22245. Logos were also generated from *Rhizorhabdus wittichii* and *W. pipientis* genomes using the same method, but with the Ssn-related motif (SRM, see definition and explanation below) query sequence, allowing no mismatches. ShinyCircos-V2.0 was used to generate the circular genome representations with the CDS, DUS, and NTS repeats mapped[60]. Secondary ssDNA structures were predicted with RNAfold (v2.6.3) using DNA parameters[61].

Search for phylogenetically conserved patterns of NTS-gene associations across Bacteria

The analysis was then expanded to examine the regions surrounding homologs of all *Nm* 8013 proteins to identify proteins spatially associated with NTS elements across bacterial genome in the NCBI reference/representative RefSeq genome database. This database is defined by

NCBI as the Entrez search term: "Bacteria[orgn] AND (reference_genome[filter] OR representative_genome[filter])". This query yielded 17718 genomes (accessed August 29, 2023, accession numbers and assembly levels in Supplementary Data 4). For each non-pseudogenized protein in the *Nm* 8013 proteome (GCA_000026965.1), each genome in this database was queried using tblastn (v2.14) to identify homologous proteins in an in-house framework using tools from the samtools (v1.17) and exonerate (v2.4.0) packages[62–64]. As the focus was to identify deeply conserved gene-NTS associations across bacteria, we excluded all genomes from *Neisseriaceae*, as defined in the NCBI taxonomy database. Conservatively, a homolog for the query protein from the *Nm* 8013 proteome in a given genome was considered to be present if a tblastn search yielded at least one alignment of *E*-value $< 1e^{-10}$ with amino acid identity of $\geq 60\%$ over 75% of the protein query's length. Only the top scoring alignment was considered for each *Nm* protein-genome pair, and so a maximum of a single homologous protein per genome was considered. Then, the group of homologous proteins identified was filtered to remove closely related homologs with $\geq 90\%$ amino acid identity using CD-HIT[65]. Using an in-house script, with tools from the samtools (v1.17), bedtools (v2.30.0), and exonerate (v2.4.0) packages[62–64], for each protein in the *Nm* 8013 proteome, genomic regions surrounding the corresponding gene of all homologous proteins discovered in the genome database were then extracted, along with their 5 kb 5′ and 3′ flanking regions (total 10 kb). Finally, these flanking regions were subjected to fuzznuc searches conducted on both strands for the NTS sequence, permitting a single additional mismatch. The results of this analysis were then summarized using a custom R (v4.3.3) script[66]: for each protein in the *Nm* 8013 proteome, the number of species containing homologs found in the genome database and the fraction of those organisms containing homologs for which at least one NTS element was found in both the 5′ and 3′ 5 kb flanking regions were tabulated and plotted.

SsnA-NTS association in bacterial genomes
To examine the relationship between NTS repeats and the *ssnA* gene, their association was then investigated at the genome level in the NCBI reference/representative database described above. For each genome, the number of NTS elements was determined using fuzznuc, searching in both strands with one further position mismatches permitted. Then, the presence or absence of a homolog of the SsnA protein was determined by using tblastn and an in-house script to query the *Nm* SsnA protein sequence (CAX49033.1) against each translated genome: a homolog was considered present if the search yielded an alignment of > 75% of the query sequence length and an E-value of $<1e^{-10}$. A stricter criterion of 60% amino acid identity over the aligned length was also used.

To confirm these results without biasing the analysis by using the SsnA protein from *N. meningitidis* as a query sequence, the analysis was repeated using the CDD database. First, psiblast was used to query with the position-specific scoring matrix (PSSM) that defines the GIY-YIG_unchar_3 protein family (cd10448) in the NCBI CDD database to find homologs of the GIY-YIG_unchar_3 subfamily, using an *E*-value of $<1e^{-10}$. Then, in a more conservative criterion, those potential GIY-YIG_unchar_3 family proteins were used as queries in a rpsblast to search of the NCBI CDD database, and a GIY-YIG_unchar_3 subfamily protein was considered present in the genome if the rpsblast score was >100.

Discovery of NTS*var* and definition of a degenerated search sequence for SRM elements
For the non-targeted search for repeated motifs associated with the *ssnA* gene, that is to say without providing a query sequence, the MEME (v5.5.1) server was used[22,23]. In the 23 *Neisseriaceae* strains identified as encoding the *ssnA* gene, the regions corresponding to 5 kb upstream and downstream of this gene were extracted then submitted to the MEME server with the any number of repetitions (anr) algorithm. Two

motifs were searched with a size constraint of 10 to 40 bp, in both DNA strands. To identify NTS_var, the five genomes (see Fig. S1c) containing no canonical NTS around *ssnA* were subjected to a similar independent analysis, this time only on the 1 kb 5′ and 3′ of the gene. The NTS_var sequence was found in *N. wadsworthii* and closely related species.

Recognizing that the sequence of NTS elements may differ at broader phylogenetic scales, a degenerated search sequence for more divergent NTS-like elements was therefore defined, which we defined as Ssn-related motifs (SRM). First, SRM elements were identified by manually inspecting the genomic regions flanking the homolog of the *Nm* SsnA protein in a diverse sample of 29 bacterial genomes (Supplementary Data 5). Then, SRM elements were manually curated, and representative sequences were aligned using MAFFT (v7.505) using the X-INS-i algorithm to incorporate structural information[58]. From this alignment, a degenerated search sequence of variable length for SRMs was manually created from the nucleotide position weight matrix: HGTCWYN(0-2)YYVSBRHN(0-5)WNVSBRGN(0-2)RNCYNB. The analyses described above searching for NTS-gene and NTS-ssnA associations, and presented in Supplementary data 3 and 4 were then repeated using this SRM search sequence.

To contextualize the number of SRM elements found in bacterial genomes and their significance, a genome shuffling experiment was conducted. For each genome, 100 shuffled genomes were generated, preserving di-nucleotide frequencies using the fasta-shuffle-letters program from the MEME suite (v5.5.5)[67]. These shuffled genomes were queried for SRM elements, and the mean SRM counts, which form a genome-specific empiric estimate of the number of expected SRM elements due to chance were then compared to the SRM counts in the original genome, yielding a fold enrichment in SRM elements for each genome.

To confirm these results, a further large-scale de novo analysis of regions flanking regions encoding homologs of SsnA was performed in the NCBI representative/reference database. Using the database of homologs of SsnA (defined above using tblastn and an *E*-value of $<1e^{-10}$), we extracted the 5 kp upstream and downstream of the region encoding the SsnA homolog in a given genome (using the method described above), and submitted the resulting sequences for all genomes with an SsnA homolog to a motif search using glam2 from the MEME suite (v5.5.0, parameters -r 15 -2 -n 1000000 -a 20 -b 45 -w 28)[68].

GIY-YIG protein superfamily and Ssn protein family phylogenies
The reference/representative bacterial genome database defined above was used to extract all proteins containing GIY-YIG domains. First, a custom NCBI conserved domain database (CDD)[36] was created using the makeprofiledb tool from the blast+ package (v2.14)[69] using the 25 family and subfamily position-specific scoring matrices (PSSMs) that comprise the cd00719 superfamily. Then using a script, rpsblast was used to query all annotated proteins for each genome in the reference/representative database against this custom CDD database. The bitscore thresholds associated with each PSSM used by rpsblast to define a hit were lowered by 10%, to minimize the risk that divergent proteins not included in the alignments used to define the CDD's PSSM definitions would be excluded. Seqkit (v2.5.1)[70] was then used to extract all proteins with a significant alignment to our custom CDD database from each predicted proteome. This resulted in a total of 46332 GIY-YIG proteins across 17718 genomes. Very long, possibly misannotated proteins (> 750aa, 369/46332 proteins) were filtered and then, for computational tractability, proteins were clustered using CD-HIT (v4.8.1)[65] using a global identity threshold of 60% (v4.8.1, options −c 0.6, -G 1), resulting in a dataset of 8528 representative proteins. An alignment was generated with MAFFT[58] (v7.505, options −retree 2 −global-pair), and a phylogenetic analysis was performed with IQ-TREE (v2.2.2.7)[71], using the standard model selection and 1000 ultrafast bootstrap replicates (-m TEST -B 1000), which determined the best fit

model was the VT + G4 model. The resulting phylogenetic tree was visualized with iTOL (v6)[72]. The Ssn protein family was first identified using a blastp search with the Nm SsnA protein sequence (CAX49033.1) as a query, against all predicted proteins of each bacterial genome, in turn (options: E-value 1×10^{-10} with a query alignment cover of 75%). Protein length was calculated by the fastalength program in the exonerate package and plotted using iTOL[72]. Using the rpsblast blast hits described above and a custom script, each protein was assigned a top CDD family based on the top rpsblast hit. Where the superfamily PSSM was the top rpsblast hit, the next best hit (by bit-score) was reported if one was found. Then, the distribution of SRMs in the 500 bp flanking the genomic region encoding each protein was also examined using a custom script, and proteins were identified as being flanked if at least one SRM was detected in both the 5′ and 3′ flanking regions. Using iTOL, the resulting phylogeny and annotations were examined, and a phylogenetic clade representing the proposed Ssn nuclease family was identified by the overlapping distributions of SsnA blastp hits, GIY-YIG unchar_3 hits, flanking SRMs and protein lengths.

To generate the Ssn family phylogeny, 22,845 protein sequences containing GIY-YIG domains were retrieved from the cd10448 family of the NCBI Conserved Domain Database (CDD). Subsequent clustering, retaining only one homologous sequence of 90% identical proteins by CD-HIT (v4.8.1 -t 0 -c 0.90 options)[65], and removing sequences longer than 150 amino acids by SeqKit[70] (v2.3.0, seq -M 150 -g options) resulted in a total of 7,051 protein sequences. These proteins were aligned by MAFFT version (v7.407)[58], and the resulting alignment was used for the subsequent phylogenetic analysis. Best-fit model and maximum-likelihood phylogeny were performed using IQ-TREE version (v2.1.2) and 1000 ultrafast bootstraps[71]. Complete taxonomic information was added to group the sequences into class and family using an in-house python script. A phylogenetic tree was constructed using iTOL server version 6.6[73].

SsnA/DNA-binding assays
Electrophoretic mobility shift assays (EMSA) were performed by mixing 0.5 μM of purified 6xHis-SsnA with 0.1 μM of 6-FAM (6-Carboxyfluorescein) 5′-labeled oligonucleotides (Sigma) in a homemade NEBuffer2.1 (NEB) lacking magnesium. The DNA fragments were labeled with a single dye molecule to minimize the effect of labeling on the protein-DNA interactions. The mix was incubated at room temperature for 30 min, then supplemented with native loading buffer (50% glycerol, 10 mM Tris-HCl pH7.5, 0.03% bromophenol blue) and loaded onto a native 10% TBE-acrylamide (19:1 solution, Fisher) gel. Electrophoreses were performed at 200 V for 40 minutes in 1x TBE running buffer. Gels were imaged on a Typhoon FLA9500 (GE Healthcare) using Cy2 detection settings, and mean band intensities over a defined area were quantified using ImageJ software[74]. Fluorescent oligonucleotides are listed in Supplementary Data 8.

Nuclease assays
Purified 6xHis-SsnA (0.5 μM) was mixed with 0.8 μM of 6-FAM 5′-labeled oligonucleotides (Sigma) in NEBuffer2.1 (NEB), and the solution was incubated at 37 °C for 1h30. Then, formamide loading buffer was added (95% formamide, 20 mM EDTA, 0.2% orange G) and the samples were boiled for 3 min prior to electrophoresis. Migration was done at 200 V for 1h15 in a 17.5 % TBE-acrylamide gel with 8 M urea, using TBE as running buffer. The Typhoon FLA9500 scanner was used to image the gels with the 473 nm laser and the BPB1 (530DF20) emission filter for 6-FAM, and the 635 nm laser and the LPR (665 LP) emission filter for Cy5, and mean band intensities over a defined area were quantified using ImageJ software version 1.52 P and 1.35t[74]. When needed, oligos were annealed by mixing equimolar amounts in annealing buffer (10 mM Tris-HCl pH8, 50 mM NaCl, 1 mM EDTA), heating them to 95 °C and letting them cool down.

For assessment of the sequence specificity of SsnA, EMSA and nuclease assays were performed as described on mutated versions of NTS75 (oligos NTS75mut), where single-nucleotide mutations were introduced along the NTS repeat. Each mutated position was substituted for any other nucleotide (A to C/G/T). Nuclease assays involving oligonucleotide hybridization with a quencher were conducted, post-hybridization (equimolar amounts in annealing buffer as above, heated to 95 °C and cooled down gradually), in 96-well microplates of f-bottom black, fluotrac, Greiner. The enzyme buffers and concentrations remained consistent with the previous experiments. Fluorescent oligonucleotides were used at a concentration of 1 μM, while non-fluorescent ones were used at varying concentrations (see legend of relevant figures). Measurements were carried out using a Victor³V 1420 multilabel counter (PerkinElmer), with P490/F535 filters for top-reading and an exposure time of 1 sec.

Rolling circle amplification (RCA) and SsnA digestion
Similarly to[75], single-stranded ssDNA oligonucleotides were hybridized at an equimolar concentration (5 μM) with short, partially complementary oligonucleotides (T7F/R) by heating at 95 °C followed by gradual cooling. These hybrids were then circularized using T4 ligase (NEB) according to the supplier's protocol. This ligation product (1 μM) was used to perform rolling circle amplification (RCA). The RCA was carried out with phi29 DNA polymerase (NEB) for 16 h at 30 °C, followed by enzyme inactivation for 10 min at 65 °C.

The nuclease assay of RCA-NTS75F was performed with one-tenth of the RCA product and 2 μM SsnA in r2.1 buffer. One-tenth of the RCA product of the HemiR sequence was hybridized by heating at 95 °C and then cooling, with 1 μM of an unlabeled oligonucleotide 39-75 F in a final volume of 10 μl. Four microliters of this hybridization solution were used for the nuclease test, following the previously described procedure. The incubation times with SsnA are indicated in the corresponding figures. The cleavage product was also visualized without fluorescence (Fig. 6a, c) using a TBE-acrylamide gel with 8 M urea that was incubated with an intercalating dye (ssGreen for ssDNA, lumiprobe) bath.

In the case of fluorescence quenching experiments (Fig. 6e), different concentrations of oligonucleotides (as indicated in the figure) were circularized and amplified as described above. Fluorescent oligonucleotides bearing a 6-FAM and a quencher (BHQ1) were used at a final concentration of 0.2 μM for hybridization with the RCA. Nuclease assays, post-hybridization, were conducted in 96-well black-bottom microplates (fluotrac, Greiner). The enzyme buffers were similar to previous experiments, except that the enzyme SsnA was used at a final concentration of 1.5 μM. Measurements were carried out with a Cytation 3 image reader from Biotek, using wavelengths of 490/520 for bottom reading and automatic gain adjustment.

Quantitative transformation assays
Overnight cultures of *N. meningitidis* were resuspended to an OD_{600} of 1.0 (2.5×10^8 CFU/ml) in GCB media supplemented with 10 mM $MgCl_2$. Then, 200 ng of linearized plasmid DNA were added to 100 μl of the bacterial suspensions, and the mixes were incubated statically at 37 °C with 5% CO_2 for 2 h. Serial dilutions were performed, and the appropriate dilutions were plated on selective media to quantify the transformants, as well as on non-selective GCB plates to quantify the total CFUs. Transformation rates represent the ratio of resistant CFUs on total CFUs. For transformation assays using genomic DNA, *Nm* strains were first transformed with the appropriate plasmid. Then, whole genomic DNA (gDNA) from the resulting transformants was purified using QIAamp mini kits (Qiagen), and 200 ng was used for quantitative transformation assays as described above. Figure 4a: Transformation rates of *N. meningitidis* 8013 and mutants using pKOYebN::Sp, b, *N. meningitidis* 8013 and mutants using pCompInd c, of *Nm* 8013 *rpsL*K43R *yebN*::MC58 and mutants using p5′3′inCm, d, *Nm* 8013

*rpsL*K43R NTSmut and mutants using p5'3'inCmMut2. Fig. S5a: Transformation rates of *Nm* 8013 and mutants using pKOopaA::Cm, b, of *Nm* 8013 and mutants using pKONMV_0504::Cm, **c**, of *Nm* 8013 and mutants using gDNA of *Nm* 8013 rpsLK43R *yebN*::Cm d, of *Nm* 8013 and mutants using pKOlgtFrfaK::Cm, e, of *Nm* 8013 and mutants using gDNA of *Nm* *rpsL*K43R f, of *Nm* 8013 and mutants using gDNA of *Nm* 8013 NalR.

For co-culture transformation experiments, recipient *Nm* strains were generated to have an inactive chloramphenicol resistance gene in place of the *yebN* gene, while the donor *Nm* and *N. musculi* strains carried the functional gene, differing only by a non-sense mutation. To ensure unidirectional exchange of DNA from the donors to the recipients thereby minimizing false positives, the recipient strains also carried a kanamycin resistance gene elsewhere in the genome. This way, exchange of the Cm^R allele through a single recombination event is exponentially more frequent than the transfer of the entire Km^R cassette from the recipient strains to the donor strains, which would require double-homologous recombination of the regions flanking the cassette. Donor and recipient strains were resuspended to an OD_{600} of 0.1 in GCB media supplemented with 10 mM $MgCl_2$. Fifty microliters of each strain were mixed together in a 96-well plate, and the co-cultures were incubated statically 3 h at 37 °C with 5% CO_2, before plating appropriate dilutions onto GCB plates with both kanamycin and chloramphenicol. Figure 4e, DNA exchange between *Nm* 8013 *rpsL*[K43R] yebN::CmStop co-cultured with *Nm* 8013 *rpsL*[K43R] yebN::Cm donor strain. f, as in e, but where the donor is *N. musculi* DSM 101846 yebNOut::Cm.

Statistical analyses and reproducibility
All statistical analyses were performed with GraphPad Prism 7.00 from at least three independent experiments, unless otherwise indicated. One- or two-way ANOVAs were performed with a Tukey's correction for multiple comparison when more than two datasets were compared, while two-tailed Mann–Whitney tests were performed to compare two datasets. Mean was used in all the graphs. All the experiments, measures, and gels presented here were performed at least three times or more with similar results, except for Fig. 6d, e, which were conducted twice with independent purifications.

Reporting summary
Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability
The raw and intermediate data files, along with explanatory notes, are included in the GitHub data supplement BactSymEvol/Chenaletal-Discovery-of-Ssn[34] and source data file[76] [https://doi.org/10.5281/zenodo.14623444 and https://doi.org/10.5281/zenodo.14630455], which is structured to match the bioinformatic analyses presented in this article. The reference/representative bacterial genome database is not included due to archive size constraints but consists of publicly available data from the NCBI genome database. A full list of NCBI genome accession numbers is provided in the data supplement. Accession numbers for bacterial genomes sequenced in our laboratory and used to construct the Neisseriaceae genome database are available in Supplementary Data 1. The newly sequenced genomic data generated in this study have been deposited in the NCBI Bioproject database under accession code PRJNA647881.

Code availability
Custom scripts, in combination with publicly available or published software packages, were used for data collection and analysis in this manuscript. All scripts are available in the GitHub data supplement BactSymEvol/Chenaletal-Discovery-of-Ssn[34] and source data file[76] [https://doi.org/10.5281/zenodo.14623444 and https://doi.org/10.5281/zenodo.14630455].

References
1. Gaj, T., Gersbach, C. A. & Barbas, C. F. 3rd. ZFN, TALEN, and CRISPR/Cas-based methods for genome engineering. *Trends Biotechnol.* **31**, 397–405 (2013).
2. Knott, G. J. & Doudna, J. A. CRISPR-Cas guides the future of genetic engineering. *Science* **361**, 866–869 (2018).
3. Dunin-Horkawicz, S., Feder, M. & Bujnicki, J. M. Phylogenomic analysis of the GIY-YIG nuclease superfamily. *BMC Genomics* **7**, 98 (2006).
4. Kleinstiver, B. P. et al. The I-TevI nuclease and linker domains contribute to the specificity of monomeric TALENS. *G3* **4**, 1155–1165 (2014).
5. Kleinstiver, B. P. et al. Monomeric site-specific nucleases for genome editing. *Proc. Natl Acad. Sci. USA* **109**, 8061–8066 (2012).
6. Beurdeley, M. et al. Compact designer TALENs for efficient genome engineering. *Nat. Commun.* **4**, 1762 (2013).
7. Catlin, B. & Cunningham, L. Transforming activities and base contents of deoxyribo-nucleate preparations from various Neisseriae. *Microbiology* **26**, 303–312 (1961).
8. Berry, J.-L., Cehovin, A., McDowell, M. A., Lea, S. M. & Pelicic, V. Functional analysis of the interdependence between DNA Uptake sequence and its cognate ComP receptor during natural transformation in Neisseria species. *PLOS Genet* **9**, e1004014 (2013).
9. Frye, S. A., Nilsen, M., Tønjum, T. & Ambur, O. H. Dialects of the DNA uptake sequence in Neisseriaceae. *PLoS Genet.* **9**, e1003458 (2013).
10. Domingues, S. et al. Natural transformation facilitates transfer of transposons, integrons and gene cassettes between bacterial species. *PLoS Pathog.* **8**, e1002837–e1002837 (2012).
11. Koomey, M. Competence for natural transformation in Neisseria gonorrhoeae: a model system for studies of horizontal gene transfer. *APMIS Suppl.* **84**, 56–61 (1998).
12. Frosch, M. & Meyer, T. F. Transformation-mediated exchange of virulence determinants by co-cultivation of pathogenic Neisseriae. *FEMS Microbiol. Lett.* **100**, 345–349 (1992).
13. Joseph, B. et al. Virulence evolution of the human pathogen neisseria meningitidis by recombination in the core and accessory genome. *PLoS ONE* **6**, e18441 (2011).
14. Marri, P. R. et al. Genome sequencing reveals widespread virulence gene exchange among human Neisseria species. *PLoS ONE* **5**, e11835 (2010).
15. Lerminiaux, N. A. & Cameron, A. D. Horizontal transfer of antibiotic resistance genes in clinical environments. *Can. J. Microbiol.* **65**, 34–44 (2019).
16. Domingues, S., Nielsen, K. M. & da Silva, G. J. Various pathways leading to the acquisition of antibiotic resistance by natural transformation. *Mob. Genet. Elem.* **2**, 257–260 (2012).
17. Spratt, B. G., Bowler, L. D., Zhang, Q. Y., Zhou, J. & Smith, J. M. Role of interspecies transfer of chromosomal genes in the evolution of penicillin resistance in pathogenic and commensal Neisseria species. *J. Mol. Evol.* **34**, 115–125 (1992).
18. Rotman, E. & Seifert, H. S. The genetics of Neisseria species. *Annu. Rev. Genet.* **48**, 405–431 (2014).
19. Blokesch, M. Natural competence for transformation. *Curr. Biol.* **26**, R1126–R1130 (2016).
20. Johnston, C., Martin, B., Fichant, G., Polard, P. & Claverys, J.-P. Bacterial transformation: distribution, shared mechanisms and divergent control. *Nat. Rev. Microbiol.* **12**, 181–196 (2014).
21. Sparling, P. F. Genetic transformation of Neisseria gonorrhoeae to streptomycin resistance. *J. Bacteriol.* **92**, 1364–1371 (1966).
22. Biswas, G., Sox, T., Blackman, E. & Sparling, P. Factors affecting genetic transformation of Neisseria gonorrhoeae. *J. Bacteriol.* **129**, 983–992 (1977).
23. Dillard, J. P. Genetic manipulation of Neisseria gonorrhoeae. *Curr. Protoc. Microbiol.* **23**, 4A. 2.1–4A. 2.24 (2011).

24. Duffin, P. M. & Seifert, H. S. DNA uptake sequence-mediated enhancement of transformation in Neisseria gonorrhoeae is strain dependent. *J. Bacteriol.* **192**, 4436–4444 (2010).

25. Ambur, O. H., Frye, S. A. & Tønjum, T. New functional identity for the DNA uptake sequence in transformation and its presence in transcriptional terminators. *J. Bacteriol.* **189**, 2077–2085 (2007).

26. Sisco, K. L. & Smith, H. O. Sequence-specific DNA uptake in Haemophilus transformation. *Proc. Natl Acad. Sci. USA* **76**, 972–976 (1979).

27. Goodman, S. D. & Scocca, J. J. Identification and arrangement of the DNA sequence recognized in specific transformation of Neisseria gonorrhoeae. *Proc. Natl Acad. Sci. USA* **85**, 6982–6986 (1988).

28. Schoen, C. et al. Whole-genome comparison of disease and carriage strains provides insights into virulence evolution in Neisseria meningitidis. *Proc. Natl Acad. Sci. USA* **105**, 3473–3478 (2008).

29. Bille, E. et al. A chromosomally integrated bacteriophage in invasive meningococci. *J. Exp. Med.* **201**, 1905–1913 (2005).

30. Kawai, M., Uchiyama, I. & Kobayashi, I. Genome comparison in silico in Neisseria suggests integration of filamentous bacteriophages by their own transposase. *DNA Res.* **12**, 389–401 (2005).

31. Bentley, S. D. et al. Meningococcal genetic variation mechanisms viewed through comparative analysis of serogroup C strain FAM18. *PLoS Genet.* **3**, e23 (2007).

32. Haas, R. & Meyer, T. F. The repertoire of silent pilus genes in Neisseria gonorrhoeae: evidence for gene conversion. *Cell* **44**, 107–115 (1986).

33. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–410 (1990).

34. Veyrier, F. et al. Discovery of the widespread site-specific single-stranded nuclease family Ssn. https://doi.org/10.5281/zenodo. 14623445 (2025).

35. Dubnau, D. DNA uptake in bacteria. *Annu. Rev. Microbiol.* **53**, 217–244 (1999).

36. Wang, J. et al. The conserved domain database in 2023. *Nucleic Acids Res.* **51**, D384–D388 (2023).

37. Veyrier, F. J. & M., C. Endonucleases that selectively cleave single-stranded nucleic acids and uses thereof. PCT/CA2022/ 051668 (2022).

38. Graver, B. A., Chakravarty, N. & Solomon, K. V. Prokaryotic Argonautes for in vivo biotechnology and molecular diagnostics. *Trends Biotechnol.* **42**, 61–73 (2024).

39. Makarova, K. S., Wolf, Y. I., van der Oost, J. & Koonin, E. V. Prokaryotic homologs of Argonaute proteins are predicted to function as key components of a novel system of defense against mobile genetic elements. *Biol. Direct* **4**, 29 (2009).

40. Jinek, M. et al. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* **337**, 816–821 (2012).

41. Budroni, S. et al. Neisseria meningitidis is structured in clades associated with restriction modification systems that modulate homologous recombination. *Proc. Natl Acad. Sci. USA* **108**, 4494–4499 (2011).

42. Parkhill, J. et al. Complete DNA sequence of a serogroup A strain of Neisseria meningitidis Z2491. *Nature* **404**, 502–506 (2000).

43. Cehovin, A. et al. Specific DNA recognition mediated by a type IV pilin. *Proc. Natl Acad. Sci. USA* **110**, 3065–3070 (2013).

44. Berry, J. L., Cehovin, A., McDowell, M. A., Lea, S. M. & Pelicic, V. Functional analysis of the interdependence between DNA uptake sequence and its cognate ComP receptor during natural transformation in Neisseria species. *PLoS Genet.* **9**, e1004014 (2013).

45. Davidsen, T. et al. Biased distribution of DNA uptake sequences towards genome maintenance genes. *Nucleic Acids Res.* **32**, 1050–1058 (2004).

46. Kim, W. J. et al. Commensal Neisseria Kill Neisseria gonorrhoeae through a DNA-Dependent Mechanism. *Cell Host Microbe* **26**, 228–239 e228 (2019).

47. Kline, K. A., Sechman, E. V., Skaar, E. P. & Seifert, H. S. Recombination, repair and replication in the pathogenic Neisseriae: the 3 R's of molecular genetics of two human-specific bacterial pathogens. *Mol. Microbiol.* **50**, 3–13 (2003).

48. Davidsen, T., Tuven, H. K., Bjoras, M., Rodland, E. A. & Tonjum, T. Genetic interactions of DNA repair pathways in the pathogen *Neisseria meningitidis*. *J. Bacteriol.* **189**, 5728–5737 (2007).

49. Hamilton, H. L. & Dillard, J. P. Natural transformation of Neisseria gonorrhoeae: from DNA donation to homologous recombination. *Mol. Microbiol.* **59**, 376–385 (2006).

50. Ronning, D. R. et al. Active Site Sharing and Subterminal Hairpin Recognition in a New Class of DNA Transposases. *Mol. Cell* **20**, 143–154 (2005).

51. He, S. et al. The IS200/IS605 Family and "Peel and Paste" Single-strand Transposition Mechanism. *Microbiol. Spectr.* **3**, https://doi. org/10.1128/microbiolspec.MDNA3-0039-2014 (2015).

52. Nunvar, J., Huckova, T. & Licha, I. Identification and characterization of repetitive extragenic palindromes (REP)-associated tyrosine transposases: implications for REP evolution and dynamics in bacterial genomes. *BMC Genomics* **11**, 44 (2010).

53. Ton-Hoang, B. et al. Structuring the bacterial genome: Y1-transposases associated with REP-BIME sequences. *Nucleic Acids Res.* **40**, 3596–3609 (2012).

54. Fukui, K. et al. The GIY-YIG endonuclease domain of Arabidopsis MutS homolog 1 specifically binds to branched DNA structures. *FEBS Lett.* **592**, 4066–4077 (2018).

55. Dunin-Horkawicz, S., Feder, M. & Bujnicki, J. M. Phylogenomic analysis of the GIY-YIG nuclease superfamily. *BMC Genomics* **7**, 1–19 (2006).

56. Desai, N. A. & Shankar, V. Single-strand-specific nucleases. *FEMS Microbiol. Rev.* **26**, 457–491 (2003).

57. Rice, P., Longden, I. & Bleasby, A. EMBOSS: the European molecular biology open software suite. *Trends Genet.* **16**, 276–277 (2000).

58. Katoh, K. & Standley, D. M. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* **30**, 772–780 (2013).

59. Crooks, G. E., Hon, G., Chandonia, J.-M. & Brenner, S. E. WebLogo: a sequence logo generator. *Genome Res.* **14**, 1188–1190 (2004).

60. Wang, Y. et al. shinyCircos-V2. 0: leveraging the creation of Circos plot with enhanced usability and advanced features. *iMeta* **2** e109 (2023).

61. Hofacker, I. L. Vienna RNA secondary structure server. *Nucleic Acids Res.* **31**, 3429–3431 (2003).

62. Slater, G. S. C. & Birney, E. Automated generation of heuristics for biological sequence comparison. *BMC Bioinform.* **6**, 1–11 (2005).

63. Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842 (2010).

64. Danecek, P. et al. Twelve years of SAMtools and BCFtools. *Gigascience* **10**, giab008 (2021).

65. Li, W. & Godzik, A. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* **22**, 1658–1659 (2006).

66. R Development Core Team. *R: A Language and Environment for Statistical Computing* (R Foundation for Statistical computing, 2010).

67. Bailey, T. L., Johnson, J., Grant, C. E. & Noble, W. S. The MEME Suite. *Nucleic Acids Res.* **43**, W39–W49 (2015).

68. Frith, M. C., Saunders, N. F., Kobe, B. & Bailey, T. L. Discovering sequence motifs with arbitrary insertions and deletions. *PLoS Comput Biol.* **4**, e1000071 (2008).

69. Camacho, C. et al. BLAST+: architecture and applications. *BMC Bioinform.* **10**, 421 (2009).

70. Shen, W., Sipos, B. & Zhao, L. SeqKit2: a Swiss army knife for sequence and alignment processing. *Imeta* **3**, e191 (2024).

71. Nguyen, L. T., Schmidt, H. A., von Haeseler, A. & Minh, B. Q. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.* **32**, 268–274 (2015).

72. Letunic, I. & Bork, P. Interactive Tree of Life (iTOL) v6: recent updates to the phylogenetic tree display and annotation tool. *Nucleic Acids Res.* **52**, W78–W82 (2024).

73. Letunic, I. & Bork, P. Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. *Nucleic Acids Res.* **49**, W293–W296 (2021).

74. Schneider, C. A., Rasband, W. S. & Eliceiri, K. W. NIH Image to ImageJ: 25 years of image analysis. *Nat. Methods* **9**, 671–675 (2012).

75. Schmidt, T. L. et al. Scalable amplification of strand subsets from chip-synthesized oligonucleotide libraries. *Nat. Commun.* **6**, 8634 (2015).

76. Veyrier, F. et al. Discovery of the widespread site-specific single-stranded nuclease family Ssn-sourcedata. https://doi.org/10.5281/zenodo.14630455 (2025).

Author contributions

M.C. and A.R.M. did most experiments, visualization and formal analysis. L.B.H. did most of the bioinformatic experiments. A.R.M, L.B.H. M.C. wrote and revised the manuscript. A.S.K., C.N., E.B., M.B., J.P., and M.E. did some experiments and formal analysis. F.J.V. did some experiments, formal analysis, conceptualized and supervised the work, acquired funding, provided resources, wrote and revised the manuscript.

Competing interests

M.C. and F.J.V. declare the following competing interests: co-inventors on the international patent application no. PCT/CA2022/051668 related to this work. The remaining authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41467-025-57514-1.

Correspondence and requests for materials should be addressed to Frédéric J. Veyrier.

Reprints and permissions information is available at http://www.nature.com/reprints

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.