nature portfolio

Peer Review File

Discovery of the Widespread Site-Specific Single-Stranded Nuclease Family Ssn

Corresponding Author: Dr Frédéric Veyrier

This file contains all reviewer reports in order by version, followed by all author rebuttals in order by version.

Version 0:

Reviewer comments:

Reviewer #1

(Remarks to the Author)
The manuscript by Chenal and colleagues describes modifications of the known DNA repeat dRS3, herein called NTS and NTS-like, and single-domain enzymes from the GIY-YIG endonucleases that cleave ssDNA at NTS and NTS-like sites, herein named SsnA, SsnB, and SsnC. These are the first known site-specific and ssDNA-specific nucleases and the paper shows that these nucleases are present in many phyla of bacteria, especially in symbionts of plants and animals. The paper thoroughly explores the binding and cutting site requirements for one of the SsnA proteins. The paper shows that NTS and NTS-like sites often flank variable genes, especially those involved in pathogenesis, and shows that the Ssn/NTS system restricts incoming ssDNA and lowers recombination rates at these genes. This result is interpreted as a restriction barrier specifically for genes flanked by NTS and NTS-like sites. The paper also demonstrates some activities related to biotechnology applications and discusses some biotech implications.
Overall, this paper is highly impactful for at least three reasons: (i) it is the first demonstration of a site-specific ssDNA endonuclease; (ii) it assigns this function to a large number of hypothetical genes in a wide variety of bacteria and proposes a common function to limit recombination of specific flanked genes; and (iii) it provides an early blueprint for potential biotech functions focused on site-specific ssDNA detection and cleavage.
The manuscript covers a lot of ground and is very thorough for an initial discovery. The experiments are motivated logically and generally well-designed and implemented.
We do have some specific suggestions to improve the manuscript, described below.

Conceptual/systematic
The evolutionary work should be improved or simplified. Given the experimental strengths of the paper, simplification would be ok, but certainly if the authors want to support their current claims about evolution, the first point would need to be followed to provide clear results.
- The manuscript doesn't clearly show a relationship between Ssn proteins and other members of the GIY-YIG superfamily and doesn't carefully establish the monophyly of these enzymes and therefore their categorization as a single family within the superfamily. The paper should have a phylogenetic analysis and/or CLANs analysis based on GIY-YIG domains to determine whether these single-domain GIY-YIG proteins are monophyletic are not. If so, their classification into one family is justified and this would suggest that the single-domain enzymes evolved once from a more complex multidomain ancestor. If not (i.e., if the single-domain GIY-YIG enzymes form many groups) then these enzymes may be the ancestors of multiple groups of multiple domain enzymes and lumping the single-domain enzymes into one family would not be justified. Either result would be exciting, but an attempt should be made to resolve the two scenarios. In the paper, a single family is justified in lines 158-160 and a reference is made to Fig. S1d-e but these analyses definitely do not resolve the two scenarios.
- Contrary to text in the manuscript, the relationships between SsnA, SsnB, and SsnC are not clear. The paper states that the CLANs analysis and the phylogenetic analysis are concordant, but they are not. The CLANs analysis clearly resolves SsnC from SsnA/SsnB, but the resolution of SsnA from SsnB is not clear. We are not familiar with the details of CLANs, but the paper needs to either present evidence that these two groups are resolved and statistically supported or this division should be removed. Similarly, in the phylogenetic analysis, SsnA makes up the majority of the phylogeny and it is not monophyletic. In the phylogeny, it appears that SsnC is a recently evolved group of homologs that are found only in some FCB genomes. This group does appear to be supported by both analyses but it's a very small group and it's unclear whether it should be given equal status to SsnA in particular. In general, enzymes should be classified into groups that aren't supported by evolutionary history.

- This is a minor point but it looks like the vast majority of Ssn homologs are found in the kingdom Pseudomonadati (see here https://www.microbiologyresearch.org/docserver/fulltext/ijsem/74/1/ijsem006242.pdf?expires=1722113810&id=id&accname=guest&checksum=67786B62DDC41983EC1BCC7C0BE328F7) with only a few present in Bacillati, which generally follows the Gm+/Gm- ring in Figure 5A. Do you agree? Maybe this suggests their evolution after the divergence of these two kingdoms?

The paper is not consistent regarding the relationship between NTS sites and recombination rates, and therefore the biological role of the NTS/Ssn system is not totally clear. In particular, the paper cites previous papers in the introduction and discussion to support statements that dRS3 repeats are at recombination hotspots, but the results shown in Figure 4 clearly show that the NTS/Ssn system decreases recombination rates. How can these conflicting statements be resolved? Is the implication that genes flanked by NTS sequences might be very small fragments of DNA within recombination hotspots that are protected from recombination? What is the relationship between NTS sites and DUS sites? More generally, it superficially seems incongruent that variable genes (rather than core genes) are guarded against recombination. Could this phenomenon be framed within a context of selfish DNA?

In general, the text and figure related to biotech applications is not as well communicated as previous parts of the manuscript. This part of the paper is important and improvements would help communicate the results and implications. Some specifics:

- Lines 285-287. The antecedent for “this” is unclear and it’s not clear what four homologs the text is referring to here. Site-specific needs a hyphen.
- Figure 6 in general, consider text headings within the figure panels to mention the major products/applications. Figure 6b and the text should be improved to clarify the constructs and results. Please consider a slightly expanded description of the use of the NTS/Ssn system for programmable cleavage.
- Application (3) is described less clearly than the other two applications. It seems that this application would be to develop molecular beacons based on engineered DNA containing NTS sequences to quantify activities that convert dsDNA to ssDNA. Is that right? Consider explaining more clearly.

In general, figures and figure legends can be improved for clarity and communication. Specific examples are below.

- Figure 1b – there are four columns with red gradients and none are clearly defined. Frequency? What does that mean? This looks like a ratio of the two elements converted to a percent. Also, consider adding a star for the clade containing N. meningiditis and N. gonorrhoeae for emphasis or color those branches or something.
- Figure 1c – there is a second scale in the center rings that is not shown.
- Figure 2b – please use the same color red as previously used in Fig 1c and include a key for the colors on b. There is space for that. Improve the font for the phylum key.
- Figure 2d – somewhere these proteins should be identified unambiguously. Domain names are difficult to read and should be clarified in a key. Presumably multimeric structure is predicted, not demonstrated?
- For all gel figures, one or more nucleotide ladders or at least an indication of size should be included.
- Figure 4, NTS sites are now red bowties where each arrowhead in the bowtie represents a half of the palindrome right? Show a legend.
- Figure 6 legend is too long for Nat. Comms rules (350 words).

The main manuscript does not contain the required methods section and the paper needs much better communication of specific experiments and results. The journal will provide its own advice, but below are some that should be congruent with the journal.

- A brief methods section should be in the main manuscript. Please check Instructions for Authors.
- Exact p-values should be used in the text, figures, and/or figure legends.
- Sequences and substitutions of all oligos should be clearly declared in supplemental tables. In some cases, there’s no information in the main figures about the key differences in DNA sequences used in experiments (e.g., Figure 3b does state key differences between NTS1-5, etc.).
- A data file is strongly recommended.

Since the NTS system is not restricted to Neisseria, please consider giving it a more general name. Note that Ssn has a general and descriptive name. This might be less confusing to the community going forward. Related to this, are NTS, NTSvar, and NTS-like sequences found in archaea or eukaryotes? It would be worth a remark.

Minor/specific

- Line 36, recently the International Committee on Systematics of Prokaryotes established kingdoms formally as the highest rank below domain in both bacteria and archaea. As such, you should definitely not use “kingdom” here. You can use “domain”. (See here https://www.microbiologyresearch.org/docserver/fulltext/ijsem/74/1/ijsem006242.pdf?expires=1722113810&id=id&accname=guest&checksum=67786B62DDC41983EC1BCC7C0BE328F7)
- Line 63, change to “genus Neisseria”
- Line 67, remove “known to be”. (In general, the paper was efficient but consider rereading carefully before submission to remove other inefficient phrases.)
- Line 72, add a comma after “Streptococcus spp.”
- Line 77-79, this sentence isn’t totally logical as written. Assuming vertical evolution, one could expect the opposite – high conservation of competency. For this sentence to make sense you need to mention the concept that this pattern suggests dynamic evolution of factors responsible for competency.
- Line 80, The sentence starts with “these bacteria” but the rest of the sentence is only focused on Neisseria,. This discrepancy should be resolved.
- Line 92, line 95, line 125, and line 129, please consider being more informative than “numerous bacteria” or “multiple bacteria”. These systems are widely distributed in many phyla, which is exciting and can be easily communicated here.
- Line 94, please consider a better description than “strikingly associated”.
- Line 125, BLAST.
- Line 129, Suggest moving nonrestrictive clause to end of sentence subheading for better flow

- Line 132-136, please introduce and clarify the logic better here. For example, consider “flanked” rather than “surrounded by”, but more importantly the preceding text doesn’t set the stage for this hypothesis. Also, the phrase “or autoregulated” needs to be explained/contextualized or dropped here.
- Line 133 and elsewhere, per Nature requirements, all scripts need to be shared on github (or somewhere else).
- Line 134, please briefly mention the methods here – what distance from Nm genes? What e-value or bitscore cutoffs?
- Line 145 and 151, here you are referring to the gene, right? If so, the notation should be ssnA. Please check the manuscript. In line 151 if you are referring to protein homology, then the protein is fine but since you are referring to the genome you should say “encode a homolog”.
- Line 147, not capital for “NTS-like”.
- Line 154, change to “codes for” or “encodes”.
- Line 156-158, is there a citation or two that shows that these extensions give these endonucleases binding specificity? If not, please qualify this.
- Line 172, the paper doesn’t previously mention a predicted cofactor binding site, so use the indirect object “a” instead of the direct object “the”.
- What is the relationship between NTSvar and NTS-like sequences? Please clarify.
- Line 201, consider mentioning some key or unusual results in a few words, per your judgement.
- Line 224, the introductory phrase “interestingly” seems overused in the results, especially because some of these results are logical and not that remarkable. It’s up to you.
- Line 292, what dye?
- Supplemental Table S8 – Recommend renaming “NTS100, NTS37, NTS28” as “NTS_100bp, NTS_37bp, NTS_28bp” to make it clear that these are the different length oligos from Figure 3c.

Reviewed by Brian Hedlund and student Ryan Doss

Reviewer #2

(Remarks to the Author)
The manuscript titled "Discovery of the Widespread Site-Specific Single-Stranded Nuclease Family Ssn" utilizes extensive sequence and biochemical analyses to investigate the unique Neisseria Transformation Sequence (NTS) elements and related site-specific single-stranded nucleases (Ssn). The NTS sequence can form stem-loop structures and is conserved in pathogenic Neisseria genomes and many bacteria. It belongs to the shorter dRS3 sequence and is crucial for integrating the host genome and facilitating horizontal gene transfer. SsnA is a small nuclease belonging to the GIY-YIG nuclease superfamily, which includes many restriction enzymes, homing endonucleases, and transposases involved in significant biological processes. Typically, members of the GIY-YIG nuclease superfamily are multi-domain proteins, with the GIY-YIG nuclease domain serving as the catalytic center and other domains controlling substrate specificity. Unlike other members of this superfamily, SsnA is a small nuclease with only the GIY-YIG nuclease domain, yet it exhibits high structural selectivity, preferring to process ssDNA with stem-loop structures like the NTS sequence. The nuclease activity of SsnA can cleave transforming DNA containing NTS repeats, thereby modulating its integration into the host genome. According to the author’s hypothesis, SsnA acts as a negative regulator of donor DNA recombination containing the NTS sequence (Figures 4g and 4h). This discovery is significant for understanding bacterial horizontal gene transfer and drug resistance, presenting novel insights that would interest the audience of Nature Communications.However, the quality of this manuscript can be improved, and some questions still need to be addressed. Major points to consider are:

1. Figures 4g and 4h suggest that donor DNA with NTS repeats is digested by SsnA, making it difficult to recombine into the host genomic DNA or resulting in a lower recombination rate. Both NTS and SsnA act as inhibitors of bacterial horizontal gene transfer. This hypothesis is questionable, as horizontal transfer is beneficial for bacteria, allowing them to acquire new DNA and adapt to environmental stress. Why would bacteria have conserved NLS and SsnA, two negative regulators, to control genome dynamics? The author should address this critical issue using reliable references and sound inferences.

2. The biochemical assay used to understand the catalytic properties is disorganized and overly complex. Most experiments need to be redesigned. First, to determine a nuclease's substrate preference, a concentration-dependent experiment is necessary; a fixed concentration is insufficient. In nuclease activity and binding assays, it is crucial to know the concentration difference of SsnA when distinguishing between preferred and non-preferred substrates. This information is essential for understanding whether the specificity of the newly identified nuclease is high or low.

Second, SsnA can digest stem-loop DNA with single-stranded 3’- and 5’-overhangs. The author should determine the specific lengths of the 3’- and 5’-overhangs required for SsnA processing, as the experiment in Figure 2c is too rudimentary. Similarly, the length of the stem region also needs to be confirmed using designed substrates. Additionally, the author should provide experiments to verify that the NTS and NTS iP (inversion of the palindrome sequence; Figure S2g) form stem-loop structures, as these structures are the foundation of all experiments.

Third, the experiments depicted in Figures 3d, S2e, and S2f are challenging to comprehend due to an overly simplistic experimental description. What do the symbols B, V, D, and H represent in the NTS75mut sequence (Table S8)? Additionally, the author needs to verify that nucleotide exchanges do not alter the stem-loop structures of the substrates. Confirming the presence of the stem-loop structure is crucial for understanding the sequence's impact on the nuclease and binding activity of SsnA.

Fourth, the product generated after SsnA cleavage is ssDNA with a 5’-OH group (lines 197-203). This differs from other

members of the GIY-YIG nuclease superfamily, such as UvrC, which produce ssDNA with 5'-phosphate and 3'-OH ends. Generally, enzymes using a similar domain (the GIY-YIG nuclease domain) yield similar products. The author should provide a reasonable explanation for this difference, supported by reliable references.

3. The author establishes that SsnA is a small protein, comprising fewer than 100 amino acids, yet exhibits both structural and some sequence preferences for the NLS sequence. This characteristic is quite remarkable. Given its size, the interaction region of SsnA with its target sequence must be small. How can a nuclease exhibit both structural and sequence preferences with such a limited interaction region? The author should provide plausible explanations or molecular mechanisms to support their hypothesis, emphasizing insights from the perspective of structural biology.

Minor points.

1. Please provide the full name of dRS3.

2. The sequence of NTS is CGTCATTCCCGCGMAVGCGGGAATCYRG. What are the symbols B, V, D, and H? This information should be shown in the table description of Table S8.

3. At line 158, "Despite sharing the characteristic 3D fold of the GIY-YIG domain, SsnA's amino acid sequence significantly diverges from other families, supporting its placement in a standalone family (Fig. S1d-e)." The author should explain how diverse Ssn is in detail and by some quantitative description, such as sequence similarity and identity.

4. At line 158, "Despite sharing the characteristic 3D fold of the GIY-YIG domain, SsnA's amino acid sequence significantly diverges from other families, supporting its placement in a standalone family (Fig. S1d-e)." The author used the Phyre2 server to infer the 3D structure of SsnA. To strengthen the analysis, the author should provide a quantitative description, such as structural alignment and RMSD values, to demonstrate that SsnA shares structural similarities with other members of the GIY-YIG nuclease family. Additionally, the author can superimpose the active site structures of SsnA and other family members to show their similarity. Highlighting the mutated sites within the structure would also make the rationale behind these mutations clearer and more comprehensible for readers.

5. The author should check if SsnA is a monomer or oligomer. This information is an important biochemical characteristic and should listed in Fig 2D, similar to other members of the GIY-YIG nuclease superfamily.

Reviewer #3

(Remarks to the Author)
I co-reviewed this manuscript with one of the reviewers who provided the listed reports. This is part of the Nature Communications initiative to facilitate training in peer review and to provide appropriate recognition for Early Career Researchers who co-review manuscripts.

Version 1:

Reviewer comments:

Reviewer #1

(Remarks to the Author)
This is Reviewer 1 from the first review, with a support review. We are very happy with the authors' responses to our criticisms. There are just a few minor issues:

- Figure 4: A two-tailed T-test depends on the data being normally distributed (parametric), but there is no test for normality. You can test for normality and then apply this statistic (if normal) or apply a non-parametric analysis – possibly Mann-Whitney.
- Line: 131: Missing SRM in the heading
- Lines 146-149: This sentence is jumbled and needs to be rewritten
- Please specify Cy2 detection settings

Review by Brian Hedlund

Reviewer #2

(Remarks to the Author)
1. Although Table S10 is not essential for the conclusion of the article, it is still essential for establishing the background of this enzyme to the reader. Thus, I still recommend keeping Table S10 in this article.
2. Tables S9 and S10 do not appear in the main article and SUPPLEMENTARY FIGURES. The author should mark Tables S9 and S10 in the main text.
3. After addressing the upper issue, All questions I raised are addressed. This article is suitable for publishing.

Reviewer #3

(Remarks to the Author)
I co-reviewed this manuscript with one of the reviewers who provided the listed reports. This is part of the Nature Communications initiative to facilitate training in peer review and to provide appropriate recognition for Early Career Researchers who co-review manuscripts.

Dear Reviewers,

We would like to thank the reviewers for their valuable time and their thoughtful comments that we believe have significantly improved the manuscript and the work in general. We think we have answered most of the reviewers' comments and you will find below a point-by-point response in bold typeface. As the manuscript is already dense, we have elected not to include all the supplementary analyses that we have performed in our revision in the article itself, but we provide such data in this response to reviewers document and its supplements. If reviewers would prefer for us to include some of it in the main document, we will be happy to do so.

Please note page and line numbers refer to the PDF version of the main text document with "Track changes" enabled.

REVIEWER COMMENTS

Reviewer #1 (Remarks to the Author):

The manuscript by Chenal and colleagues describes modifications of the known DNA repeat dRS3, herein called NTS and NTS-like, and single-domain enzymes from the GIY-YIG endonucleases that cleave ssDNA at NTS and NTS-like sites, herein named SsnA, SsnB, and SsnC. These are the first known site-specific and ssDNA-specific nucleases and the paper shows that these nucleases are present in many phyla of bacteria, especially in symbionts of plants and animals. The paper thoroughly explores the binding and cutting site requirements for one of the SsnA proteins. The paper shows that NTS and NTS-like sites often flank variable genes, especially those involved in pathogenesis, and shows that the Ssn/NTS system restricts incoming ssDNA and lowers recombination rates at these genes. This result is interpreted as a restriction barrier specifically for genes flanked by NTS and NTS-like sites. The paper also demonstrates some activities related to biotechnology applications and discusses some biotech implications.
Overall, this paper is highly impactful for at least three reasons: (i) it is the first demonstration of a site-specific ssDNA endonuclease; (ii) it assigns this function to a large number of hypothetical genes in a wide variety of bacteria and proposes a common function to limit recombination of specific flanked genes; and (iii) it provides an early blueprint for potential biotech functions focused on site-specific ssDNA detection and cleavage.
The manuscript covers a lot of ground and is very thorough for an initial discovery. The experiments are motivated logically and generally well-designed and implemented.
We do have some specific suggestions to improve the manuscript, described below.

Thanks for these encouraging comments.

Conceptual/systematic
The evolutionary work should be improved or simplified. Given the experimental strengths of the paper, simplification would be ok, but certainly if the authors want to support their current claims about evolution, the first point would need to be followed to provide clear results.
- The manuscript doesn't clearly show a relationship between Ssn proteins and other members of the GIY-YIG superfamily and doesn't carefully establish the monophyly of these enzymes and therefore their categorization as a single family within the superfamily. The paper should have a phylogenetic analysis and/or CLANs analysis based on GIY-YIG domains to determine whether these single-domain GIY-YIG proteins are monophyletic are not.

Thank you for this comment, we agree that the work as previously presented was insufficient

to support our claim that the Ssn enzymes consisted of one protein family. In an effort to answer the reviewers' very legitimate comments, we have conducted further analysis: we have estimated a maximum likelihood phylogenetic tree of all GIY-YIG containing proteins extracted from our database of bacterial reference/representative genomes. These were extracted based on a rpsblast search against a custom NCBI CDD database for the different GIY-YIG family definitions (position-specific scoring matrices, PSSMs). This analysis is presented as new Figure 5a, discussed in the text (page 9, lines 356-374), and the methods are included in the updated methods section (page 18, lines 732-764) and in the new data supplement (https://github.com/BactSymEvol/Chenaletal-Discovery-of-Ssn).

To highlight Ssn homologs in this tree, we used a blastp search with the characterized enzyme SsnA found in *N. meningiditis* (Nm) as a query and defined a list of Ssn homlogs among bacterial genomes. We then examined their position on the phylogenetic tree of clustered representatives of all GIY-YIG containing proteins. Please refer to the methods section for details of the tree construction and analysis (page 18, lines 732-764, datafiles in the github repository). The distribution of these homologs highlights a monophyletic clade of representative GIY-YIG proteins that consists of 544/553 proteins containing a GIY-YIG unchar 3 domain. This clade is also enriched in proteins flanked by SRM (previously NTS-like) elements. We therefore define this clade as the Ssn family, which is essentially synonymous with the CDD family GIY-YIG_unchar_3, particularly given the caveats of possible protein misannotation, assembly errors, phylogenetic error and/or domain fusions/transfers which may lead to rare annotation of an unchar_3 domain elsewhere in the phylogeny.

If so, their classification into one family is justified and this would suggest that the single-domain enzymes evolved once from a more complex multidomain ancestor. If not (i.e., if the single-domain GIY-YIG enzymes form many groups) then these enzymes may be the ancestors of multiple groups of multiple domain enzymes and lumping the single-domain enzymes into one family would not be justified. Either result would be exciting, but an attempt should be made to resolve the two scenarios. In the paper, a single family is justified in lines 158-160 and a reference is made to Fig. S1d-e but these analyses definitely do not resolve the two scenarios.

As suggested by the reviewer, further detailed analysis of the evolution of the Ssn family is beyond the scope of this article, but we believe this analysis is sufficient to define the Ssn family phylogenetically within the GIY-YIG superfamily, synonymous with the CDD family GIY-YIG_unchar_3 (while few exceptions may exist due to wrong family assignation or domain fusions). The text has been modified to incorporate this new analysis (page 9, lines 356-374). Development of PSSM/HMM model-based definitions of Ssn proteins will be the subject of a follow-up article. Preliminary observation of the distribution of protein length among the Ssn family (Figure 5a) and the GIY-YIG tree as a whole suggests that these proteins have evolved from a short, ~100aa ancestor. Ssn is recovered as a sister clade to proteins containing SLX1-like domains (per NCBI CDD hits; structure-specific nucleases), which are also sparsely characterized in bacteria. GIY-YIG phylogeny is challenging as the protein in many families is only 100AA long. We have proven here the function of several homologues distributed over the GIY-YIG_unchar_3 phylogeny and propose the family classification supported by a common origin and identical function (site-specific ssDNA nuclease). If the reviewers agree, we would prefer not to go further in the description of this evolutionary scenario pending more detailed analyses and follow-up publications.

- This is a minor point but it looks like the vast majority of Ssn homologs are found in the kingdom Pseudomonadati (see here https://www.microbiologyresearch.org/docserver/fulltext/ijsem/74/1/ijsem006242.pdf?expires=1722113810&id=id&accname=guest&checksum=67786B62DDC41983EC1BCC7C0BE328F7) with only a few present in Bacillati, which generally follows the Gm+/Gm- ring in Figure 5A. Do you agree? Maybe this suggests their evolution after the divergence of these two kingdoms?

We hesitate to draw detailed quantitative conclusions concerning the taxonomic distribution of homologs in this analysis (Fig. 5b), which is based on all proteins in the CDD database annotated to contain GIY-YIG_unchar_3 domains, followed by clustering at a 90% identity threshold and a length of less than 150aa. The intention of the analysis presented in Fig 5b was to show that this protein family is broadly distributed in bacteria. We do agree that Pseudomonadati does seem to dominate the distribution, and that this phylogeny could certainly be used to support the hypothesis that the origin of this protein is in Pseudomonadati with subsequently horizontal transfer to other clades of bacteria, followed by vertical evolution, but over this scale of phylogeny, a complex, reticulated evolutionary history must be assumed.

Related to this, are NTS, NTSvar, and NTS-like sequences found in archaea or eukaryotes? It would be worth a remark.

Thank you for the suggestion to examine the presence of NTS-like (now renamed SRM) sequences in Archaea and Eukaryota. We downloaded a representative/reference genome database for both Archaea and Eukaryota from the NCBI genome database on Aug 21, 2024 (as had previously been performed for Bacteria). As these genomes are often much longer than bacteria, and the SRM search motif is degenerated, it was expected to find some SRM by random chance. Given its variable length and nucleotide degeneracy, it is not at all trivial to derive the number of SRM/NTS-like elements that would be expected by random chance in a given genome. To estimate this quantity empirically and better contextualize both this analysis and our original SRM search in bacteria, we have performed genome sequence shuffling experiments as described below and in the revised text.

For Archaean genomes (n=678), we found the presence of SRM (NTS-like) motifs, but at lower numbers, and no clear blastp-defined homologs of the SsnA protein. For each genome, 100 shuffled versions of the genome were generated, preserving di-nucleotide frequencies (using MEME-suite's fasta-shuffle-letters –k 2 option), and then SRM elements were enumerated as before using fuzznuc. SRM counts in the genome and the mean of SRM counts in the 100 shuffled genomes were then normalized to counts per Mbp. Thus, observed SRM counts were then compared to the mean of counts found in shuffled genomes. Bacterial genomes shown to contain functional Ssn proteins and SRM elements (e.g. *Wolbachia pipienitis* and *Neisseria meningitidis)* have many fold enrichment in observed SRM elements (183x and 55x, respectively) relative to the corresponding randomly shuffled genomes. In contrast, Archaean genomes were observed to have at most 2-3x (at very low SRM counts). Given the lack of clear Ssn or GIY-YIG_unchar3 proteins in Archaean genomes, we believe it is unlikely that the Ssn-SRM system is present.

With respect to Eukaryota, we assembled a database of (n=952) genomes spanning eukaryotic diversity. We also detected SRM elements, in some cases in thousands of copies. We detected two cases of blastp homologs to *N. meningitidis* SsnA in the genomes of the insects *Drosophila ananassae*, and *Maniola hyperantus.* We believe these to be host genome sequence contamination derived from *Wolbachia* endosymbionts (which our analyses have demonstrated contain Ssn-SRM systems). Our conclusion is based on blastn searches of the identified genomic region encoding the SsnA homolog in each insect genome against the NCBI core nt database, for which the top hits are to *Wolbachia* sequences. We repeated the genome shuffling analysis described above but performed only 10 dinucleotide shuffles per genome (for computational tractability given

the size of the eukaryotic genomes). Compared to the shuffled genomes, the recovered SRM counts are mostly in the 1-2x range relative to shuffled genome counts. Several genomes did have higher ratios in 10-12x range (e.g. *Pieris napi*, *Biomphalaria glabrata*). Given the lack of SsnA protein homologs or unchar_3 proteins and relatively little enrichment in SRM elements, we also suspect that the Ssn-SRM system may not be broadly present in Eukaryotes.

Although we suspect that this system is very rare (or even absent) in archaeans and eukaryotes, we would prefer to reserve final conclusions for a more detailed investigation. Each genome could need individual investigation to check for putative contamination (such as with *Wolbachia*). Details of the Archaea and Eukaryota analyses are available to the reviewers in the github data supplement (direct links: https://github.com/BactSymEvol/Chenaletal-Discovery-of-Ssn/blob/main/7.Misc/Archaean_NTS_ssn_analysis.xlsx, https://github.com/BactSymEvol/Chenaletal-Discovery-of-Ssn/blob/main/7.Misc/Eukaryotes_NTS_ssn_analysis.xlsx).

- Contrary to text in the manuscript, the relationships between SsnA, SsnB, and SsnC are not clear. The paper states that the CLANs analysis and the phylogenetic analysis are concordant, but they are not. The CLANs analysis clearly resolves SsnC from SsnA/SsnB, but the resolution of SsnA from SsnB is not clear. We are not familiar with the details of CLANs, but the paper needs to either present evidence that these two groups are resolved and statistically supported or this division should be removed. Similarly, in the phylogenetic analysis, SsnA makes up the majority of the phylogeny and it is not monophyletic. In the phylogeny, it appears that SsnC is a recently evolved group of homologs that are found only in some FCB genomes. This group does appear to be supported by both analyses but it's a very small group and it's unclear whether it should be given equal status to SsnA in particular. In general, enzymes should be classified into groups that aren't supported by evolutionary history.

Thank you for this suggestion and these comments. Following the reviewers' suggestions, we have simplified the presentation of the evolutionary analysis of the Ssn family. We present a phylogenetic tree of the Ssn family based on the analysis of GIY-YIG_unchar_3 proteins (Fig. 5b) but do not attempt further subclassification of enzyme sub-families. We will reserve this important work for a follow-up article. We have modified the text to remove reference to hypothetical Ssn subfamilies (page 9-10, lines 374-375).

The paper is not consistent regarding the relationship between NTS sites and recombination rates, and therefore the biological role of the NTS/Ssn system is not totally clear. In particular, the paper cites previous papers in the introduction and discussion to support statements that dRS3 repeats are at recombination hotspots, but the results shown in Figure 4 clearly show that the NTS/Ssn system decreases recombination rates. How can these conflicting statements be resolved? Is the implication that genes flanked by NTS sequences might be very small fragments of DNA within recombination hotspots that are protected from recombination? What is the relationship between NTS sites and DUS sites? More generally, it superficially seems incongruent that variable genes (rather than core genes) are guarded against recombination. Could this phenomenon be framed within a context of selfish DNA?

Thank you for pointing us this lack of clarity in our discussion of the results. We have rephrased some of our interpretations to increase the clarity of our conclusions, and to limit our conclusions to those best supported by these data. First, we may have incorrectly stated that these NTS (or dRS3) elements have been associated with recombination hotspots. In fact, it has been shown that dRS3-adjacent genes have increased sequence diversity, explained by the fact that these repeats could be used as homologous regions for recombination (text modified page 12, lines 494-499). We have removed the term "recombination hotspot" that can imply HGT and gain of accessory genes. Consequently, it is not accessory genes that we wanted to refer to here but more genes with variation

in their sequence (such as the famous – and well maintained - *pilE* in *Neisseria*). A phenomenon referred to as antigenic variation, which can be due to mutations, exogenous DNA acquisition but also sometime recombination with endogenous DNA (ex *pilS* cassette) in *Neisseria*. These variations are under the selective pressure of hosts' immunity (as there are often antigens at the surface of bacteria). We have rephrased our discussion trying to be clearer and to be more measured in our interpretations (page 13, lines 520-544). The aim of this article is to describe the Ssn family composed of single strand and sequence-specific nucleases. We think the discovery of NTS/Ssn in *Neisseria* may be just the beginning of the characterization of a new system regulating competence that will require future refinement by our lab and other research groups.

In general, the text and figure related to biotech applications is not as well communicated as previous parts of the manuscript. This part of the paper is important and improvements would help communicate the results and implications. Some specifics:

- Lines 285-287. The antecedent for "this" is unclear and it's not clear what four homologs the text is referring to here. Site-specific needs a hyphen.
Corrected (page 10, line 399)

- Figure 6 in general, consider text headings within the figure panels to mention the major products/applications. Figure 6b and the text should be improved to clarify the constructs and results. Please consider a slightly expanded description of the use of the NTS/Ssn system for programmable cleavage.
Thank you for this comment, we were limited by the maximum length of the article. We have added some sentences to slightly expand the description (page 11, lines 441-447).

- Application (3) is described less clearly than the other two applications. It seems that this application would be to develop molecular beacons based on engineered DNA containing NTS sequences to quantify activities that convert dsDNA to ssDNA. Is that right? Consider explaining more clearly.
Thank for pointing us this lack of clarity, again we were limited by the maximum length. Application (3) explains the use of the NTS/Ssn system for the detection of DNA carrying specific sequences (currently one half of the NTS palindrome). To clarify this, we have modified the text accordingly (page 11-12, lines 456-487)

In general, figures and figure legends can be improved for clarity and communication. Specific examples are below.

- Figure 1b – there are four columns with red gradients, and none are clearly defined. Frequency? What does that mean? This looks like a ratio of the two elements converted to a percent. Also, consider adding a star for the clade containing N. meningiditis and N. gonorrhoeae for emphasis or color those branches or something.
Thank you for this comment, we had omitted the scales for the color gradients. We have also now highlighted pathogenic *Neisseria*.

- Figure 1c – there is a second scale in the center rings that is not shown.
The schematic representations of the different loci are not drawn to scale, as they are meant to illustrate a "zoom" on the parts of the genome containing a local accumulation of NTS. The legend has been modified to clarify this point.

- Figure 2b – please use the same color red as previously used in Fig 1c and include a key for the colors on b. There is space for that. Improve the font for the phylum key.
NTS are in red in the two figures. To improve clarity, we added a key for the colors.

- Figure 2d – somewhere these proteins should be identified unambiguously. Domain names are difficult to read and should be clarified in a key. Presumably multimeric structure is predicted, not demonstrated?
Given the new analysis of GIY-YIG-containing proteins presented in Fig. 5a, this analysis was both overly simplified and redundant, and so has been removed from the article.

- For all gel figures, one or more nucleotide ladders or at least an indication of size should be included.
The various gels in this study contain different ladders/markers. To maintain consistency and clarity in the figures, we have added size indications on gels that previously lacked. The ladders are visible in the original, unmodified gel images, which are provided in the raw data file with the article.

- Figure 4, NTS sites are now red bowties where each arrowhead in the bowtie represents a half of the palindrome right? Show a legend.
Thank you for pointing out this lack of clarity. No, each NTS is illustrated with a red arrow, which also indicates its direction. This has been clarified in the legend (Pages 27, line 991-992).

- Figure 6 legend is too long for Nat. Comms rules (350 words).
Corrected (Page 29, lines 1053-1078)

The main manuscript does not contain the required methods section and the paper needs much better communication of specific experiments and results. The journal will provide its own advice, but below are some that should be congruent with the journal.
- A brief methods section should be in the main manuscript. Please check Instructions for Authors.
Thank you, we were limited by the previous journal format. We have now included this section in the main manuscript (pages 15-22).

- Exact p-values should be used in the text, figures, and/or figure legends.
The detailed parameters of all statistical tests, including the degrees of freedom and exact p-values, are available in the raw data file. We have added the exact p-values in the figures.

- Sequences and substitutions of all oligos should be clearly declared in supplemental tables. In some cases, there's no information in the main figures about the key differences in DNA sequences used in experiments (e.g., Figure 3b does state key differences between NTS1-5, etc.).
Thank you for this suggestion, we have made it more clear in the supplementary tables, for example in Table S8, we have highlighted the substitutions in each oligonucleotide in red. We have also added their predicted structure as per the reviewer two's recommendations.

- A data file is strongly recommended.
We have created a github repository for the data files to complement the supplementary materials. This includes all scripts and the raw data sets used in this article (https://github.com/BactSymEvol/Chenaletal-Discovery-of-Ssn)

Since the NTS system is not restricted to Neisseria, please consider giving it a more general name. Note that Ssn has a general and descriptive name. This might be less confusing to the community going forward.
Thank you for this suggestion, given that NTS and NTSvar elements are primarily restricted to *Neisseriaceae* where they were discovered, we have elected to maintain this name for the original NTS sequence motif. For the NTS-like sequence motif, we agree with the reviewer's suggestion, and have elected to call the degenerated search motif the "Ssn-related motif (SRM)", previously "NTS-like" (Lines 199,201-203,205,370,386-387,546,589,641,696,708,710,713,716,717,720-723,758,760,764)

Minor/specific
- Line 36, recently the International Committee on Systematics of Prokaryotes established kingdoms formally as the highest rank below domain in both bacteria and archaea. As such, you should definitely not use "kingdom" here. You can use "domain". (See here https://www.microbiologyresearch.org/docserver/fulltext/ijsem/74/1/ijsem006242.pdf?expires=1722113810&id=id&accname=guest&checksum=67786B62DDC41983EC1BCC7C0BE328F7)
Thank you for this, we updated our taxonomic nomenclature accordingly (page 2-3, line 35,64).

- Line 63, change to "genus Neisseria"
Corrected (page 3, line 65)

- Line 67, remove "known to be". (In general, the paper was efficient but consider rereading carefully before submission to remove other inefficient phrases.)
Corrected (page 3, line 69).

- Line 72, add a comma after "Streptococcus spp."
Corrected (page 3, line 74).

- Line 77-79, this sentence isn't totally logical as written. Assuming vertical evolution, one could expect the opposite – high conservation of competency. For this sentence to make sense you need to mention the concept that this pattern suggests dynamic evolution of factors responsible for competency.
Sorry, one word was missing. We intended to state that 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 (page 4, line 85)

- Line 80, The sentence starts with "these bacteria" but the rest of the sentence is only focused on Neisseria,. This discrepancy should be resolved.
The rest of the sentence also concerns *Haemophilus* spp. which also have DUS sequences.

- Line 92, line 95, line 125, and line 129, please consider being more informative than "numerous bacteria" or "multiple bacteria". These systems are widely distributed in many phyla, which is exciting and can be easily communicated here.
Thank you for these suggestions, we have clarified as follows:
Line 92: referring to NTS / *Neisseriaceae)* (page 4, line 100)
Line 95: referring to the distribution of SRM/Ssn spatial relationships, which was demonstrated across bacteria (Supp. Table S3, page 4, lines 103-104).
Line 125: here we are referring to initial investigations that demonstrated SRM/NTS-like elements were widespread, further details on the distribution are given later in the article (page 5, line 148).
Line 129: In this section, we are referring to both NTS and SRM/NTS-like elements present in multiple bacterial genomes (page 5, line 152).

- Line 94, please consider a better description than "strikingly associated".
We have replaced "strikingly associated" by "spatially associated" (page 5, line 157).

- Line 125, BLAST.
Corrected (page 5, line 148).

- Line 129, Suggest moving nonrestrictive clause to end of sentence subheading for better flow
We have modified to "NTS-like motifs are associated with a specific gene (*ssnA*) in multiple bacterial genomes" (page 5, line 152).

- Line 132-136, please introduce and clarify the logic better here. For example, consider "flanked" rather than "surrounded by", but more importantly the preceding text doesn't set the stage for this hypothesis. Also, the phrase "or autoregulated" needs to be explained/contextualized or dropped here.
We have replaced the full sentence by : "1) mostly in bacterial genomes harboring NTS and 2) be spatially associated with NTS elements (for example to be possibly transcriptionally regulated by them)" (page 5, lines 157-158) . Our first hypothesis was that this sequence was serving as binding site for a putative regulator. This has allowed us to identify SsnA (even if it is not a transcriptional regulator).

- Line 133 and elsewhere, per Nature requirements, all scripts need to be shared on github (or somewhere else).
We have prepared a github repository with all scripts and bioinformatic data files. (https://github.com/BactSymEvol/Chenaletal-Discovery-of-Ssn)

- Line 134, please briefly mention the methods here – what distance from Nm genes? What e-value or bitscore cutoffs?
The details of the methods are included in the new methods section. Briefly: the SsnA (*Nm*) protein was used as a query and blasted against each genome independently with tblastn, using an alignment query coverage of 75% and E-value of $1x10^{-10}$. Then flanking regions of 5000bp were considered around each alignment satisfying those criteria. (Page 17, Lines 685-698)

- Line 145 and 151, here you are referring to the gene, right? If so, the notation should be ssnA. Please check the manuscript. In line 151 if you are referring to protein homology, then the protein is fine but since you are referring to the genome you should say "encode a homolog".
Thank you for pointing this out. Yes, we are referring to the genome encoding homologs of the protein. We have corrected the terminology. (Page 6, Line 204)

- Line 147, not capital for "NTS-like".
Done (Page 6, Line 199)

- Line 154, change to "codes for" or "encodes".
Done (Page 6, Line 210)

- Line 156-158, is there a citation or two that shows that these extensions give these endonucleases binding specificity? If not, please qualify this.
This has been reviewed in Dunin-Horkawicz, S., Feder, M. & Bujnicki, J. M. Phylogenomic analysis of the GIY-YIG nuclease superfamily. *BMC Genomics* 7, 98 (2006). We have changed binding specificity by "their functional specificities" which is more accurate. (Page 6, Lines 214-215)

- Line 172, the paper doesn't previously mention a predicted cofactor binding site, so use the indirect object "a" instead of the direct object "the". **Done (Page 7, Line 247)**

- What is the relationship between NTSvar and NTS-like sequences? Please clarify.
The NTS_{var} sequence is the SRM/NTS-like element motif that was found in the genome of *Neisseria wadsworthii* and five related *Neisseriacae* genomes through *de novo* MEME suite analysis (Figure S1c, *Neisseria dumasiana, N. canis*, *N. wadsworthii*, *N dentiae*, and *Uruburuella suis*). This has been clarified in the text by making reference to Figure S1c. The sequence motifs themselves are described in the updated methods sections. (Page 17, Lines 696-705)

- Line 201, consider mentioning some key or unusual results in a few words, per your judgement. **Done, we have added cofactor, temperature and Kd in the text. (Page 8, Lines 294-296)**

- Line 224, the introductory phrase "interestingly" seems overused in the results, especially because some of these results are logical and not that remarkable. It's up to you.
Done Some instances of "interestingly" have been removed or replaced to avoid excessive use. (Pages 7, Lines 255,272,319)

- Line 292, what dye?
This information is present in the legend of Figure 6. We use a specific dye for ssDNA and RNA : (ssGreen – Lumiprobe)

- Supplemental Table S8 – Recommend renaming "NTS100, NTS37, NTS28" as "NTS_100bp, NTS_37bp, NTS_28bp" to make it clear that these are the different length oligos from Figure 3c.
Done, we added lengths of this different oligos in table S8.

Reviewer #2 (Remarks to the Author):

The manuscript titled "Discovery of the Widespread Site-Specific Single-Stranded Nuclease Family Ssn" utilizes extensive sequence and biochemical analyses to investigate the unique Neisseria Transformation Sequence (NTS) elements and related site-specific single-stranded nucleases (Ssn). The NTS sequence can form stem-loop structures and is conserved in pathogenic Neisseria genomes and many bacteria. It belongs to the shorter dRS3 sequence and is crucial for integrating the host genome and facilitating horizontal gene transfer. SsnA is a small nuclease belonging to the GIY-YIG nuclease superfamily, which includes many restriction enzymes, homing endonucleases, and transposases involved in significant biological processes. Typically, members of the GIY-YIG nuclease superfamily are multi-domain proteins, with the GIY-YIG nuclease domain serving as the catalytic center and other domains controlling substrate specificity. Unlike other members of this superfamily, SsnA is a small nuclease with only the GIY-YIG nuclease domain, yet it exhibits high structural selectivity, preferring to process ssDNA with stem-loop structures like the NTS sequence. The nuclease activity of SsnA can cleave transforming DNA containing NTS repeats, thereby modulating its integration into the host genome. According to the author's hypothesis, SsnA acts as a negative regulator of donor DNA recombination containing the NTS sequence (Figures 4g and 4h). This discovery is significant for understanding bacterial horizontal gene transfer and drug resistance, presenting novel insights that would interest the audience of Nature Communications. However, the quality of this manuscript can be improved, and some questions still need to be addressed. Major points to consider are:

1. Figures 4g and 4h suggest that donor DNA with NTS repeats is digested by SsnA, making it difficult to recombine into the host genomic DNA or resulting in a lower recombination rate. Both NTS and SsnA act as inhibitors of bacterial horizontal gene transfer. This hypothesis is questionable, as horizontal transfer is beneficial for bacteria, allowing them to acquire new DNA and adapt to environmental stress. Why would bacteria have conserved NLS and SsnA, two negative regulators, to control genome dynamics? The author should address this critical issue using reliable references and sound inferences.

Thank you for pointing us this lack of clarity. We rephrased the discussion to be clearer and more nuanced in our affirmations, also considering reviewer 1's suggestions (p13 Line 490-544). It is important to note the *Neisseria* specific context here: in this genus, HGT does not necessarily mean acquisition of accessory genes but can also mean acquisition of homologous DNA (without new gene acquisition). The genes that are surrounded by NTS are not accessory genes but are genes for which the presence is highly conserved both within and among some of the species, but some sequence variability occurs between isolates. We think that we may have not been precise in our discussion in the first version of the manuscript and we had inadvertently induced confusion in our description (ex: referring as recombination hotspot).

Horizontal gene transfer is not universally beneficial in all contexts. In fact, *Neisseria* species (particularly *N. meningitidis* and *N. gonorrhoeae,* that have identical NTS/Ssn systems) exhibit rapid and relatively uncontrolled genome evolution, with a significant capacity for importing and integrating exogenous DNA, especially when this DNA comes from other *Neisseria* species bearing DNA Uptake Sequences (DUS). It has been shown that acquisition of exogenous DNA from other species induce death of *N. gonorrhoeae* (see reference #2 below). Demonstrating that in certain contexts, this genomic integration can be detrimental. The NTS/SsnA system could serve as a protective mechanism to limit such events. Moreover, several studies in the scientific literature highlight that horizontal gene transfer is not always advantageous for bacteria.

1. Hall, R. J., Whelan, F. J., McInerney, J. O., Ou, Y. & Domingo-Sananes, M. R. Horizontal Gene Transfer as a Source of Conflict and Cooperation in Prokaryotes. Front. Microbiol. 11, (2020).

2. Kim, W. J. et al. Commensal Neisseria Kill Neisseria gonorrhoeae through a DNA-Dependent Mechanism. Cell Host & Microbe 26, 228-239.e8 (2019).

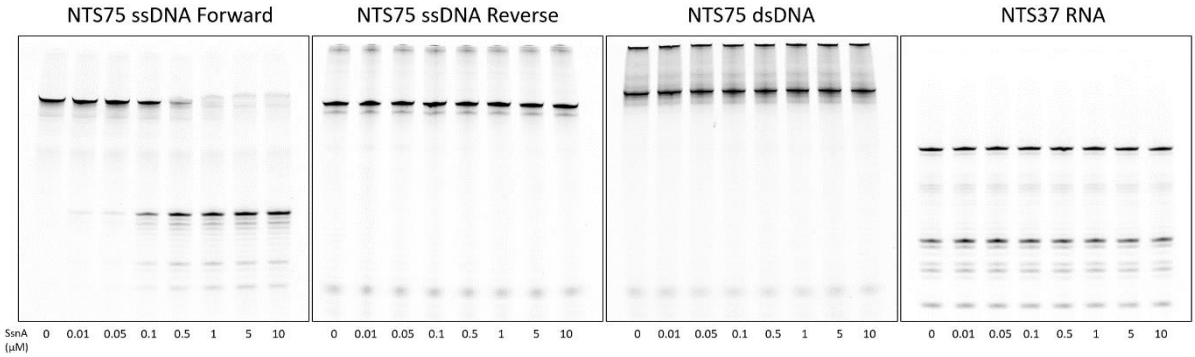
3. Acar Kirit, H., Lagator, M. & Bollback, J. P. Experimental determination of evolutionary barriers to horizontal gene transfer. BMC Microbiology 20, 326 (2020).

4. Vogan, A. A. & Higgs, P. G. The advantages and disadvantages of horizontal gene transfer and the emergence of the first species. Biology Direct 6, 1 (2011).

We have rephrased the article to be more precise and the discussion to present several alternative hypotheses.

2. The biochemical assay used to understand the catalytic properties is disorganized and overly complex. Most experiments need to be redesigned. First, to determine a nuclease's substrate preference, a concentration-dependent experiment is necessary; a fixed concentration is insufficient. In nuclease activity and binding assays, it is crucial to know the concentration difference of SsnA when distinguishing between preferred and non-preferred substrates. This information is essential for understanding whether the specificity of the newly identified nuclease is high or low.

As the reviewer pointed out, the assays presented in this article were conducted at a fixed protein concentration to identify ssDNA as the substrate for SsnA. In Figure S3, however, we varied the substrate concentrations. This approach was partially driven by the stability and solubility limitations of the SsnA protein, which constrained the range of usable concentrations in our experiments. Nevertheless, to address the reviewer's concern, we performed a new purification of the SsnA protein from *Neisseria meningitidis* and conducted nuclease assays on various substrates, varying the protein concentrations. In the article, we set the "standard" concentration of SsnA at 0.5 µM for a 1.5-hour incubation. In the new experiments presented here, we tested concentrations ranging from 0.01 µM to 10 µM, maintaining the same incubation time and substrate concentrations. Under these conditions, and with the tested concentrations, we were unable to detect significant activity on any substrate other than single-stranded DNA. We have incorporated this data in the Fig S2f and



Nuclease Activity of SsnA on Various NTS-Containing Substrates. Nuclease activity assay with different concentrations of SsnA on four substrates containing the NTS: ssDNA forward, ssDNA reverse, dsDNA, and RNA. The assays were performed under the conditions described previously, with substrate concentrations set at 0.8 µM.

mentioned it in the text (in the paragraph: "SsnA from *N. meningitidis* is a specific single-stranded nuclease that cleaves near the NTS" p7 Line 262).

Second, SsnA can digest stem-loop DNA with single-stranded 3'- and 5'-overhangs. The author should determine the specific lengths of the 3'- and 5'-overhangs required for SsnA processing, as the experiment in Figure 2c is too rudimentary.

Figure 2C presents *in silico* analyses showing conserved bases that are likely essential for the formation of the secondary structure and for cleavage by SsnA. Additionally, Figure 3C demonstrates that a length of 37 nucleotides is still permissive for cleavage, indicating that 2 nucleotides outside the palindrome on the 3' end and 5 nucleotides before the cleavage site on the 5' end are sufficient to maintain enzymatic activity. However, to refine this information and in response to the reviewer's suggestion, we performed cleavage assays with 2 nucleotides for the 5' overhang from the cleavage site (our nuclease assay's limit of detection) and varying lengths of 3' overhangs. Although less efficient, the NTS is still cut until the beginning of the stem (residue 30 in this oligo which correspond to C47 in figure Fig 3d) is no longer present. We have incorporated this data in the Fig S2e and mentioned it in the text (in the paragraph: "SsnA from *N. meningitidis* is a specific single-stranded nuclease that cleaves near the NTS" p7 Line 257 to 260).

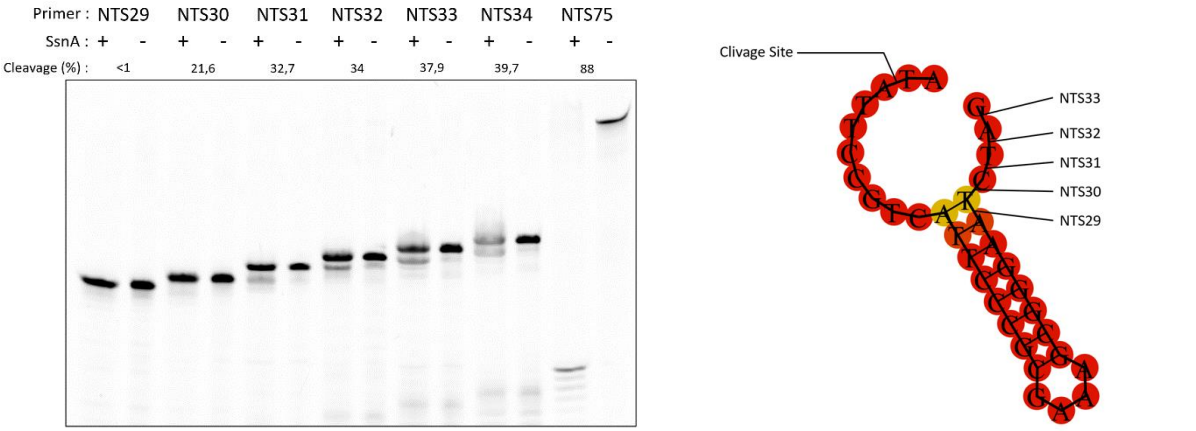


Figure Review 1: Determination of the Minimum Sequence Length Required for Nuclease Activity Around the NTS. Nuclease assay to determine the minimum sequence length necessary around the NTS for cleavage. The assay was performed under standard conditions described in the article. Primers of decreasing lengths were tested to identify the minimal sequence required for effective cleavage. The figure on the right is a schematic representation of NTS34 and the various shorter sequences used in the gel on the left.

Similarly, the length of the stem region also needs to be confirmed using designed substrates.
Regarding the length of the stem, it is already relatively short (8 nucleotides) and highly conserved throughout evolution. We have shown that the inversion of the palindrome results in a loss of function, highlighting the importance of the conserved residues in the stem. We have also inverted nucleotide from the stem and did not observe cleavage for the first two at the stem base (A29/T46 permutated to T29/A49 and T30/A47 to A30/T47). We start observing some cleavage after the third nucleotides dyad permutation (T31/A46 to A31/T46) as similarly observed in Fig. 3d. Again, this suggests that both structure and the sequence (particularly at the base of the stem) are important. We did not reduce the size of the stem as this one may no longer assemble in our condition. Importantly, in Figure 6B, we tested extending the stem while removing the loop. This experiment showed that elongating the stem increases the cleavage rate, likely due to the enhanced stability of the stem.

Additionally, the author should provide experiments to verify that the NTS and NTS iP (inversion of the palindrome sequence; Figure S2g) form stem-loop structures, as these structures are the foundation of all experiments.
Thank you for this comment, we performed structural simulations for all sequences used in this article to address this comment. These structure predictions can now be found in Table S9.

Third, the experiments depicted in Figures 3d, S2e, and S2f are challenging to comprehend due to an overly simplistic experimental description.
We have now moved the M&M in the main document (p15 Line 628). In addition, an improved explanation of the three cited figures has been added to the text (p7 Line 263-266): "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 Figure 3d, with detailed quantification provided in Figures S2g-h." We have also decided to present in the Fig. 3d, the nuclease activity without normalization with the binding activity which may led to some confusion for the readers.

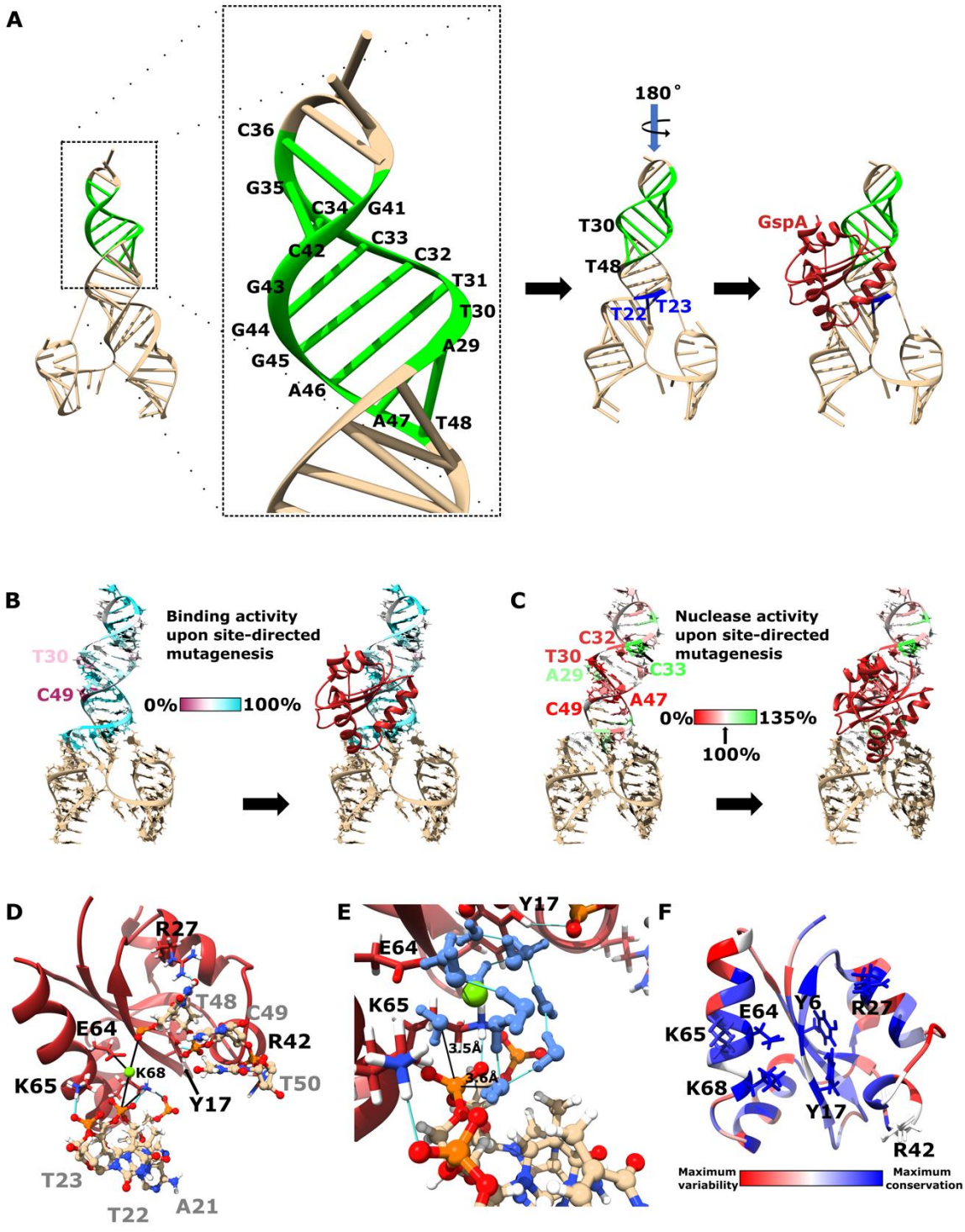
What do the symbols B, V, D, and H represent in the NTS75mut sequence (Table S8)?
These are IUPAC nucleotide degenerate base codes, for clarity, an explanatory table of degenerate bases has been added to Table S8.

Additionally, the author needs to verify that nucleotide exchanges do not alter the stem-loop structures of the substrates. Confirming the presence of the stem-loop structure is crucial for understanding the sequence's impact on the nuclease and binding activity of SsnA.
To confirm the formation of the stem-loop in the different oligonucleotides tested, we performed structure predictions under the test conditions (Mg concentration/temperature). These structures are available in Table S9

Fourth, the product generated after SsnA cleavage is ssDNA with a 5'-OH group (lines 197-203). This differs from other members of the GIY-YIG nuclease superfamily, such as UvrC, which produce ssDNA with 5'-phosphate and 3'-OH ends. Generally, enzymes using a similar domain (the GIY-YIG nuclease domain) yield similar products. The author should provide a reasonable explanation for this difference, supported by reliable references.
We agree with the reviewer that this result is somehow surprising. Nevertheless, we have repeated the experiment several times with the same result. In the last few months, we have tried several other approaches to try to definitely answer this question, but these experiments have given inconclusive results. Nevertheless, all these assays rely on the activity of enzymes (Lambda exonuclease, terminal transferase) that are supposed to be specific for 5'OH, 5'P, 3'OH and/or 3'P but some instability of the ssDNA and less strict specificity can occur in some cases, which renders the interpretation of these results difficult even for the control oligos. The result showing the 5'P and 3'OH ends was first presented as supplementary material. We think it is not essential to support the main conclusions of the article. In absence of a second method to prove our claim, we prefer to remove this figure and the sentence in the text.

3. The author establishes that SsnA is a small protein, comprising fewer than 100 amino acids, yet exhibits both structural and some sequence preferences for the NLS sequence. This characteristic is quite remarkable. Given its size, the interaction region of SsnA with its target sequence must be small. How can a nuclease exhibit both structural and sequence preferences with such a limited interaction region? The author should provide plausible explanations or molecular mechanisms to support their hypothesis, emphasizing insights from the perspective of structural biology.
We agree that this characteristic is quite remarkable, and, even without structural explanation, we think that this article which reports the discovery of the Ssn protein family as sequence specific

ssDNA endonucleases is an important work that merits publication. We are currently working in collaboration with an expert in structural biology on a follow-up study that will present robust structural data. We have performed a study to model the molecular interaction of Ssn with its NTS to guide our follow up experiments. This model created using UCSF Chimera (UCSF Chimera--a visualization system for exploratory research and analysis. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE. J Comput Chem. 2004 Oct;25(13):1605-12.), presented below, shows the possible position of Ssn on the ssDNA. It suggests also the possible involvement of Y17 of the SsnA protein which is in close contact with nucleotide T22. However, we believe these findings are too preliminary to be included in this article and will require further investigation to be publishable in a further study, particularly employing techniques including X-ray crystallography or NMR.



Molecular modeling of the complex between SsnA (from N. meningitdis) and ssDNA NTS using UCSF Chimera. A) Molecular prediction of the structure of ssDNA (length of 75 nucleotides) is shown with the stem highlighted in green, the putative cleavage site is located at nucleotides 21-22 (blue ribbon) and the SsnA protein is presented in red ribbon style. B) The results of single nucleotide site-directed mutagenesis profiles obtained by EMSA are shown on the molecular prediction of the SsnA-ssDNA complex. Residual protein binding relative to WT is shown by a color-coded gradient. C) SsnA nuclease activity on ssDNA after single site-directed mutagenesis is shown over the molecular model of the complex. D) Molecular dissection of the protein-ssDNA interface. The main atomic physical interactions, hydrogen bonds (light blue lines) and electrostatic interactions (black lines) are shown.

Residue labels use single letter coding, and protein and DNA are presented in black and gray, respectively. The atomic representation of the DNA atoms is shown with a stick and ball style while the protein side chains only use the stick style. The atoms are color coded, oxygen (red), nitrogen (blue), hydrogen (white), phosphorus (orange), carbon (light brown) and magnesium (light green). E) Atomic coordination at the putative active site with the water network is shown as blue water molecules. The distance (Å) of 2 water molecules relative to the target phosphorus atom at residue T22 is shown by black lines. (F) The degree of conservation of protein residues at the protein-DNA interface is shown. The degree of conservation was assessed by the Consurf method (Ashkenazy et al., 2010 & Celniker et al., 2013).

Minor points.

1. Please provide the full name of dRS3.
The information has been added to the introduction at the first occurrence of the abbreviation dRS3 (p4 Line 90).

2. The sequence of NTS is CGTCATTCCCGCGMAVGCGGGAATCYRG. What are the symbols B, V, D, and H? This information should be shown in the table description of Table S8.
The reading table of degenerate bases has been added to Table S8.

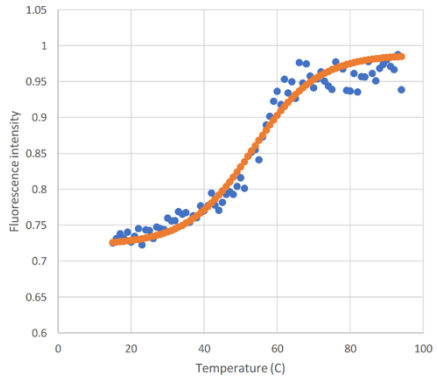
3. *At line 158, "Despite sharing the characteristic 3D fold of the GIY-YIG domain, SsnA's amino acid sequence significantly diverges from other families, supporting its placement in a standalone family (Fig. S1d-e)." The author should explain how diverse Ssn is in detail and by some quantitative description, such as sequence similarity and identity.*
Thanks for this comment that is also in line with those of Reviewer 1. We have removed part of this sentence from the article (p6 Line 213) as the classification of Ssn as a family is not based on sequence divergence from other families. It is based on common phylogenetic origin of the cd10448 and a common sequence-specific ssDNA endonuclease activity of these homologs as described here. Nevertheless, we still present improved 3D fold prediction of Ssn GIY-YIG domain to show the similarity with GIY-YIG domain of other nucleases (Fig. S1d-e).

4. At line 158, "Despite sharing the characteristic 3D fold of the GIY-YIG domain, SsnA's amino acid sequence significantly diverges from other families, supporting its placement in a standalone family (Fig. S1d-e)." The author used the Phyre2 server to infer the 3D structure of SsnA. To strengthen the analysis, the author should provide a quantitative description, such as structural alignment and RMSD values, to demonstrate that SsnA shares structural similarities with other members of the GIY-YIG nuclease family. Additionally, the author can superimpose the active site structures of SsnA and other family members to show their similarity. Highlighting the mutated sites within the structure would also make the rationale behind these mutations clearer and more comprehensible for readers.
We have followed the reviewer's advice to improve our predictive model by using the latest version of the AlphaFold server to obtain a structure prediction for the SsnA protein. In addition, to address the request for a structural comparison, we utilized the TM-align tool to compare the SsnA prediction with three GIY-YIG proteins described in the scientific literature. The structural superpositions and corresponding scores (RMSD and TM) are reported in Table S10. These data are not essential for the conclusion of the article and could be removed if the reviewer would prefer.

OBJ

5. The author should check if SsnA is a monomer or oligomer. This information is an important biochemical characteristic and should listed in Fig 2D, similar to other members of the GIY-YIG nuclease superfamily.

While this article focuses primarily on the enzymatic activity of SsnA, we have explored preliminary approaches to address whether SsnA is monomeric or oligomeric. First, we investigated the protein's state under our purification and experimental conditions. Thermal unfolding assays via fluorescence (ratio 330/350 nm) revealed a single unfolding curve, suggesting that SsnA remains monomeric under these conditions. The analysis of these data using the Boltzmann method also indicates that the denaturation temperature of this monomeric protein is 53.25 °C.
To assess potential oligomerization when interacting with ssDNA, we attempted two methods: SEC-MALS and Two-PM. However, the limited solubility of SsnA prevented us from achieving the concentrations required for SEC-MALS and the small size of the protein (11 kDa) posed a limitation for two-PM. Future work dedicated to the structural characterization of SsnA will focus on optimizing protein production and concentration conditions, which will also be essential for obtaining crystallographic structures.



Thermal Unfolding Analysis of SsnA to Determine Its Monomeric State. Thermal unfolding assays using fluorescence (330/350 nm ratio) were conducted to investigate whether SsnA is monomeric or oligomeric under purification and experimental conditions. The assays were performed with a temperature variation of 1°C per minute in SsnA's purification buffer. The results show a single unfolding curve, indicating that SsnA remains monomeric under these conditions.

Dear Reviewers,

We would like to sincerely thank the Reviewers for their valuable comments, suggestions, and guidance, which have greatly helped improve the quality of our manuscript.

Below, you will find a detailed point-by-point response to the Reviewers' comments, along with the corresponding modifications made to the manuscript.

We hope the revisions address all concerns satisfactorily, and we remain open to further suggestions if needed.

REVIEWERS' COMMENTS

Reviewer #1 (Remarks to the Author):

This is Reviewer 1 from the first review, with a support review. We are very happy with the authors' responses to our criticisms. There are just a few minor issues:

- Figure 4: A two-tailed T-test depends on the data being normally distributed (parametric), but there is no test for normality. You can test for normality and then apply this statistic (if normal) or apply a non-parametric analysis – possibly Mann-Whitney.

It is true that under these conditions, a non-parametric Mann-Whitney test seems more appropriate. We have therefore replaced the T-test with a Mann-Whitney test, and the p-values have been updated in the figures and in the Source Data file.

- Line: 131: Missing SRM in the heading

This oversight has been corrected, thank you for bringing it to our attention.

- Lines 146-149: This sentence is jumbled and needs to be rewritten

Thanks for pointing us this, we have reformulated :" Despite their presence in diverse bacteria, some *Neisseria* genomes that contained a homolog of *ssnA* did not have NTS elements. We therefore conducted 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 NTSvar (Fig. 1a-b, Fig. S1c). This prompted us to build a methodology to further capture diversity in NTS-like elements.

- Please specify Cy2 detection settings

A sentence detailing the acquisition parameters of the fluorophores has been added as follows: 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,

Reviewer #2 (Remarks to the Author):

1. Although Table S10 is not essential for the conclusion of the article, it is still essential for establishing the background of this enzyme to the reader. Thus, I still recommend keeping Table S10 in this article.

We agree with the reviewer that Table S10 is useful for general understanding, and we will therefore keep it in the article as recommended.

2. Tables S9 and S10 do not appear in the main article and SUPPLEMENTARY FIGURES. The author should mark Tables S9 and S10 in the main text.

The two supplementary tables have been cited in the main text and in the supplementary material as requested by the reviewer.

3. After addressing the upper issue, All questions I raised are addressed. This article is suitable for publishing.

Reviewer #3 (Remarks to the Author):

I co-reviewed this manuscript with one of the reviewers who provided the listed reports. This is part of the Nature Communications initiative to facilitate training in peer review and to provide appropriate recognition for Early Career Researchers who co-review manuscripts.