ABSTRACT: With the discovery of increasingly more functional noncoding RNAs (ncRNAs), it becomes eminent to more strongly consider them as important players during species evolution. Although tests for negative selection of ncRNAs already exist since the beginning of this century, the SSS-test is the first one for also investigating positive selection. When analyzing selection in ncRNAs, it should be taken into account that selection pressures can independently act on sequence and structure. We applied the SSS-test to explore the evolution of ncRNAs in primates and identified more than 100 long noncoding RNAs (IncRNAs) that might evolve under positive selection in humans. With this test, it is now possible to more thoroughly include ncRNAs into evolutionary studies.

KEYWORDS: RNA, consensus structure, structural conservation, positive selection

To understand how species evolve and adapt to their environment, tests for natural selection have been developed. The common assumption is that parts of the genome that are responsible for adaptive phenotypic changes evolve faster than other parts. Most proteins and nucleic acids exert their biological function by means of well-defined interactions. The specificity of functional interactions as well as the need to avoid undesired binding activities translates into selection pressures on both the sequence and the 3-dimensional structure of proteins and nucleic acids. A relatively simple test for estimating selection pressures on protein-coding genes has been developed in the 1980s and relates the rate of nucleotide changes that cause an amino acid change (non-synonymous changes) to the rate of silent nucleotide changes (synonymous changes), referred to as $d_{N}/d_{S}$ ratio. Ratios much smaller than 1 indicate negative selection, i.e., conservation of the protein sequence. Higher ratios are usually interpreted as relaxed constraint. If that ratio is positive, the excess of amino acid changes is compatible with accelerated evolution or a sign of positive selection. Despite the increasing acknowledgment that ncRNAs are functional, a comparable test for non-coding RNA (ncRNA) genes did not exist until recently.

Importantly, in the case of RNAs, structure-formation is dominated both thermodynamically and kinetically by the secondary structure, i.e., the pattern of base pairs and unpaired bases. The simplicity of RNA secondary structures, and their conservation from synonymous and non-synonymous substitutions in the secondary structure. This specific type of base pairing patterns are indicative of negative selection, in particular compensatory substitutions, such as the replacement of a GC pair by a CG, AU, or UA pair. On the other hand, substitutions that disrupt base pairs hint at relaxed constraints or positive selection. Conceptually, this is not different from synonymous and non-synonymous substitutions in the open reading frames (ORFs) of protein-coding genes. There is, however, an important practical difference between ORFs and RNA secondary structures: although codons are local in sequence, secondary structures are inherently nonlocal, usually involving pairs that are long-range with respect to the sequence. As a consequence, this assessment of selection pressures on secondary structure requires completely different computational tools.

It is important to realize that molecules are typically subject to multiple, superimposed selection pressures. For protein-coding genes, e.g., functional elements such as SECIS or Internal Ribosomal Entry Sites (IRES) require tightly constrained RNA secondary structures within protein-coding sequences. This specific type of superimposed selection pressures yields substitution patterns that are recognizable by specialized computational tools. Similar situations are observed in ncRNAs. For tRNAs, e.g., the clover-leaf secondary structure and the 3-dimensional
L-shape are required for loading into the ribosome and recognition for charging essentially independent of the sequence. On the other hand, tRNAs have an internal pol-III promoter, whose sequence must be maintained to ensure expression. Selection may also act on the expression level. For instance, the choice of rare codons as well as highly stable mRNA secondary structure may hamper translation. Carlini et al.\(^3\) proposed that the balance between codon bias and mRNA secondary structure is mediated through the third codon position: here, natural selection might favor high GC or AT content to increase base pairing for weakly expressed genes and the opposite for highly expressed genes. It is a nontrivial, and largely unsolved task to disentangle such superimposed selective force. Presently, available tools only model a single effect or at most a pair of specific selection pressures.

Selection pressures that independently act to maintain superimposed sequence and secondary structure features can lead to incongruent conservation of sequence and structure: in this case, sequence patterns and structural elements are shifted relative to each other. As a consequence, analogous base pairs no longer correspond to homologous sequence positions. This type of incongruent evolution violates the basic assumptions of all tools that measure secondary structure conservation: the secondary structure will not appear conserved in a sequence-based method only evaluates whether the secondary structures are less diverged than expected for the observed divergence of RNA secondary structures.\(^{14}\) As a remedy, the SSS-test associates the probability of each structural change with a structural change for both substitutions and indels. In this model, scores close to zero indicate negative selection and higher scores are indicative of positive selection. Empirical calibration suggests that scores higher than 10 are a strong indication of positive selection within the primate group.

Researchers who wish to investigate selective pressures on ncRNAs should be mindful of the biological question and choose the most suitable approach and software (Table 1), keeping in mind the different selection pressures (Figure 1).

There are several advantages to the approach taken by the SSS-test: First, it can be used for detecting signs of positive as well as negative selection. Second, it allows identifying changes in structures as well as in stability. Third, small RNAs as well as lncRNAs can be investigated; in the latter case, local structures will be tested for selection. We applied the SSS-test to more than 15,000 human lncRNAs with orthologs in various primates and identified 110 lncRNAs that are candidates for being under positive selection in humans.\(^3\) We observed two types of patterns among these candidates: Some candidates, such as LINC02217, contain local structures with completely different shapes, whereas other candidates, such as SIX3-AS1, maintain their structure but with a clearly increased stability in human compared with their orthologs. We further performed the SSS-test to investigate which lncRNAs that
have been associated with psychiatric disorders might evolve under positive selection. We discovered 8 lncRNAs that possess local structures with signs of positive selection in humans. The candidates we identified can now be further tested functionally, to decipher if and how they might be involved in human evolution, for instance, in the evolution of cognitive abilities.

The SSS-test and related software are available at https://github.com/waltercostamb/SSS-test and can now be applied for further evolutionary questions. We propose that any new genome project could annotate ncRNA genes in addition to protein-coding genes and scan for RNA structures under selection. Existing genome data and ncRNA databases could be mined and analyzed for selected ncRNAs. Biomedical studies have repeatedly found disease-associated variants within ncRNA genes. To gain further insights into the functions of such genes, their evolutionary history could be investigated with the SSS-test.

Although the SSS-test is certainly a powerful test for investigating the evolution of ncRNAs, there is still ample room for improvement: presently, the cutoffs for deeming a structure to evolve under selection are empirically determined and thus need to be calibrated by the user for each dataset. In our study, we required that the candidate structures are among the most conserved structures across the phylogeny, but demonstrate a relatively strong change in a single lineage, e.g., humans. Although the workflow can be extended to detect distinct selective pressures in different lineages, it still depends on the existence of a well-conserved ancestral structure.

In some cases, it is possible not only to identify a locus under positive selection but also to reconstruct the evolutionary history itself with some accuracy. This amounts to determining the order of substitution events and can be achieved under the assumption that the structural differences between extant and ancestral structure represent the direction of the selective force.13

Taken together, the time has come to learn more about the evolutionary history of various ncRNA genes and their role in

| SELECTIVE PRESSURE | METHOD | LEVEL OF ANALYSIS |
|--------------------|--------|-------------------|
| Positive selection  | SSS-test3 | Secondary structure |
| Accelerated evolution | Pollard et al.15 | Primary sequence |
| Negative selection  | R-scape,7 RNAz,8 cmfinder,9 grna16 Alifold2,17 EvoFold,18 SISSH19 SSS-test3 | Secondary structure |

Table 1. Types of selective pressures on noncoding RNAs and how to detect them.

Figure 1. Types of selection pressures in ncRNAs: (1) positive selection, acting on the structure, in which one species acquires a structural change in the orthologous ncRNA with an advantage over the ancestral structure; (2) accelerated evolution, acting on the primary sequence, in which the sequence of a ncRNA accumulates a relatively high number of changes compared with its orthologs over a short time span; and (3) negative selection, acting on the structure, in which the ncRNA structure is maintained across orthologs over relatively long evolutionary time.
species evolution. The SSS-test can serve to identify candidates to prioritize for further functional investigations.

**Author Contributions**

KN, MBWC, CHzS and PFS conceived and wrote the manuscript.

**ORCID iD**

Katja Nowick [https://orcid.org/0000-0003-3993-4479](https://orcid.org/0000-0003-3993-4479)

**REFERENCES**

1. Li W-H, Wu C-I, Luo C-C. A new method for estimating synonymous and nonsynonymous rates of nucleotide substitution considering the relative likelihood of nucleotide and codon changes. *Mol Biol Evol*. 1985;2:150-174.
2. Yang Z, Nielsen R. Estimating synonymous and nonsynonymous substitution rates under realistic evolutionary models. *Mol Biol Evol*. 2000;17:32-43.
3. Walter Costa MB, Honer Zu Siederdissen C, Dunjic M, Stadler PF, Nowick K. SSS-test: a novel test for detecting positive selection on RNA secondary structure. *BMC Bioinformatics*. 2019;20:151.
4. Meyer IM, Miklos I. Statistical evidence for conserved, local secondary structure in the coding regions of eukaryotic mRNAs and pre-mRNAs. *Nucleic Acids Res*. 2005;33:6338-6348.
5. Carlini DB, Chen Y, Stephan W. The relationship between third-codon position nucleotide content, codon bias, mRNA secondary structure and gene expression in the drosophilid alcohol dehydrogenase genes Adh and Adhr. *Genetica*. 2001;159:623-633.
6. Wald M, Will S, Wolfinger M, Hofacker IL., Stadler PF. Bis-alignments as models of incongruent evolution and RNA sequence and structure. In: CIBB 2019. bioArxiv:10.1101/631606.
7. Rivas E, Clements J, Eddy SR. A statistical test for conserved RNA structure shows lack of evidence for structure in lncRNAs. *Nat Methods*. 2017;14:45-48.
8. Washietl S, Hofacker IL., Stadler PF. Fast and reliable prediction of noncoding RNAs. *Proc Natl Acad Sci USA*. 2005;102:2454-2459.
9. Yao Z, Weinberg Z, Ruzzo WL. CMfinder—a covariance model based RNA motif finding algorithm. *Bioinformatics*. 2006;22:445-452.
10. Ancel LIV, Fontana W. Plasticity, evolvability, and modularity in RNA. *J Exp Zool*. 2000;288:242-283.
11. Pollard KS, Salama SR, King B, et al. Forces shaping the fastest evolving regions in the human genome. *PLoS Genet*. 2006;2:e168.
12. Pollard KS, Salama SR, Lambert N, et al. An RNA gene expressed during cortical development evolved rapidly in humans. *Nature*. 2006;443:167.
13. Walter Costa MB, Honer Zu Siederdissen C, Tulpan D, Stadler PF, Nowick K. Temporal ordering of substitutions in RNA evolution: uncovering the structural evolution of the human accelerated region 1. *J Theor Biol*. 2018;438:143-150.
14. Walter CMB, Honer zu Siederdissen C, Dunjic M, Stadler PF, Nowick K. Supplement for “SSS-test: a novel test for detecting positive selection on RNA secondary structure.” *BMC Bioinformatics*. 2019;20:151.
15. Pollard KS, Salama SR, Lambert N, et al. Supporting material for: “An RNA gene expressed during cortical development evolved rapidly in humans.” *Nature*. 2006;443:167.
16. Rivas E, Eddy SR. Noncoding RNA gene detection using comparative sequence analysis. *BMC Bioinformatics*. 2001;2:8.
17. Washietl S, Hofacker IL. Consensus folding of aligned sequences as a new measure for the detection of functional RNAs by comparative genomics. *J Mol Biol*. 2004;342:19-30.
18. Pedersen JS, Bejerano G, Stiepel A, et al. Identification and classification of conserved RNA secondary structures in the human genome. *PLoS Comput Biol*. 2006;2:e33.
19. Gesell T, Washietl S. Dinucleotide controlled null models for comparative RNA gene prediction. *BMC Bioinformatics*. 2008;9:248.