

# Selection Pressures on RNA Sequences and Structures

Katja Nowick<sup>1</sup> , Maria Beatriz Walter Costa<sup>2</sup>,  
Christian Höner zu Siederdisen<sup>3</sup> and Peter F Stadler<sup>3,4,5,6,7</sup>

<sup>1</sup>Human Biology Group, Institute for Biology, Department of Biology, Chemistry, Pharmacy, Freie Universität Berlin, Berlin, Germany. <sup>2</sup>Embrapa Agroenergia, Brasilia, Brazil. <sup>3</sup>Bioinformatics Group, Department of Computer Science, and Interdisciplinary Center for Bioinformatics, Universität Leipzig, Leipzig, Germany. <sup>4</sup>Max Planck Institute for Mathematics in the Science, Leipzig, Germany. <sup>5</sup>Department of Theoretical Chemistry, Universität Wien, Wien, Austria. <sup>6</sup>Facultad de Ciencias, Universidad Nacional de Colombia, Bogotá, Colombia. <sup>7</sup>Santa Fe Institute, Santa Fe, NM, USA.

Evolutionary Bioinformatics  
Volume 15: 1–4  
© The Author(s) 2019  
Article reuse guidelines:  
sagepub.com/journals-permissions  
DOI: 10.1177/1176934319871919



**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 (lncRNAs) 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

**RECEIVED:** July 21, 2019. **ACCEPTED:** July 29, 2019.

**TYPE:** Commentary

**FUNDING:** The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by CNPq Brasil/scholarship of Science without Borders (246039/2012-4) (MBWC), the Volkswagen Foundation within the initiative “Evolutionary Biology” (KN), the Deutsche Forschungsgemeinschaft as part of the SPP 1738 (MBWC, KN, and PFS), and in part by the German Academic Exchange Service (DAAD), proj.no. 57390771.

**DECLARATION OF CONFLICTING INTERESTS:** The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

**CORRESPONDING AUTHOR:** Katja Nowick, Human Biology Group, Institute for Biology, Department of Biology, Chemistry, Pharmacy, Freie Universität Berlin, Königin-Luise-Strasse 1-3, 14195 Berlin, Germany. Email [katja.nowick@fu-berlin.de](mailto:katja.nowick@fu-berlin.de)

**COMMENT ON:** Walter Costa MB, Höner Zu Siederdisen C, Dunjic M, Stadler PF, Nowick K. *SSS-test: a novel test for detecting positive selection on RNA secondary structure.* *BMC Bioinformatics*. 2019;20(1):151. doi:10.1186/s12859-019-2711-y. PubMed PMID: 30898084. PubMed Central PMCID: PMC6429701. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6429701/>

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<sup>1,2</sup> 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 *Ka/Ks* or *dN/dS* 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 changing Mutations 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.<sup>3</sup>

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 discrete combinatorial nature, makes it possible to describe

selection pressures acting on the structure in terms of comparably simple rules that pertain to the preservation and turnover of base pairs. Sequence variations that locally maintain 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 SElenoCystein Insertion Sequences (SECIS) or Internal Ribosomal Entry Sites (IRES) require tightly constrained RNA secondary structures within protein-coding sequences. This specific type of superimposed selective pressures yields substitution patterns that are recognizable by specialized computational tools.<sup>4</sup> 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.<sup>5</sup> 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 alignment, whereas in structure-based alignments nonhomologous nucleotides are aligned thus leading to an exaggerated estimate of compensatory base pairs. Tools to identify such cases are only in an exploratory stage of development at best.<sup>6</sup>

Over the last two decades, several methods have become available to evaluate negative/stabilizing selection of secondary structures, mostly aimed at classical structured RNAs such as tRNAs, rRNA, or snRNAs. A common assumption of all these methods is that selection acts to *preserve* individual base pairs. The difference between the strict consensus model of *R-scape*<sup>7</sup> and the reduced rate of evolution model implicit in *RNAz*<sup>8</sup> and *cmfinder*<sup>9</sup> is the strength of the selective pressure. In the strict consensus model, a conserved core structure is assumed whose base pairs are present in all homologs under consideration. Thus, one expects to observe compensatory substitutions giving direct statistical evidence for the preserved base pairs. The idea is implemented in *R-scape*.<sup>7</sup> The more relaxed model only evaluates whether the secondary structures are less diverged than expected for the observed divergence of the underlying sequences: *RNAz* therefore measures indirect evidence in the form of a “structure conservation index” that presents the ratio of folding energies between individual folds and consensus structure, and a *z*-score quantifying the folding energy relative to randomized sequences. It has been shown in Ancestral and Fontana<sup>10</sup> that stabilizing selection on the secondary structure results in more negative folding energies over evolutionary time scales.

Probabilistic models can be used to determine the type of selection acting at a given *locus*. For negative selection, the

expectation is that the rate of change is (very) low. Higher change rates can indicate either accelerated or positive evolution. Accelerated evolution is characterized by higher accumulation of changes in a short amount of time.<sup>11</sup> To identify accelerated regions, one should first identify negative selection for the orthologous locus in other species and then test for accumulation of species-specific changes.<sup>12</sup> Analyzing human accelerated regions (HARs), it seems likely that more than one evolutionary force shapes them,<sup>11</sup> including positive selection.<sup>11,13</sup>

In contrast to accelerated evolution, positive evolution occurs when the changed *locus* yields an advantage to the organism, being actively selected for throughout evolution in a longer time frame.

Although accelerated evolution is detectable at the primary sequence level alone, it is necessary to consider a phenotypic level for the detection of positive selection, to identify an advantage over the ancestral state. For ncRNAs and proteins, one should account for changes in the structure. This poses challenges for ncRNAs. Although it suffices for proteins to distinguish synonymous from non-synonymous substitutions, such a binary classification does not appear to work well for RNA secondary structures.<sup>14</sup> As a remedy, the *SSS-test* associates the probability of each structural change with a background model to calculate the likelihood of a change being merely random or being selected for. An excess of structural change indicates positive selection, whereas an excess of changes that are structure-conserving supports negative selection.<sup>3</sup> The *SSS-test* combines scoring models for 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.<sup>3</sup> 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

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

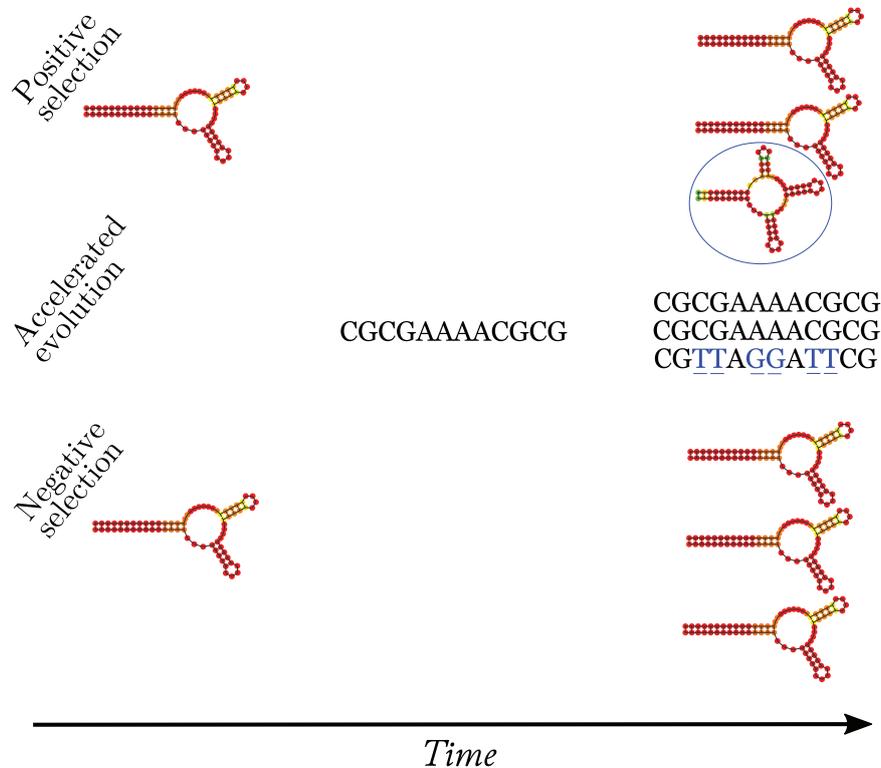
SELECTIVE PRESSURE	METHOD	LEVEL OF ANALYSIS
Positive selection	SSS-test <sup>3</sup>	Secondary structure
Accelerated evolution	Pollard et al. <sup>15</sup>	Primary sequence
Negative selection	R-scape, <sup>7</sup> RNAz, <sup>8</sup> cmfinder, <sup>9</sup> qrna <sup>16</sup> AlifoldZ, <sup>17</sup> EvoFold, <sup>18</sup> SISIz, <sup>19</sup> SSS-test <sup>3</sup>	Secondary structure

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.<sup>13</sup>

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



**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>

### REFERENCES

- 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.
- Yang Z, Nielsen R. Estimating synonymous and nonsynonymous substitution rates under realistic evolutionary models. *Mol Biol Evol.* 2000;17:32-43.
- Walter Costa MB, Honer Zu Siederdisen 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.
- 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.
- 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. *Genetics.* 2001;159:623-633.
- Waldl M, Will S, Wolfinger M, Hofacker IL, Stadler PF. Bi-alignments as models of incongruent evolution and RNA sequence and structure. In: *CIBB 2019*. bioArxiv:10.1101/631606.
- 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.
- Washietl S, Hofacker IL, Stadler PF. Fast and reliable prediction of noncoding RNAs. *Proc Natl Acad Sci U S A.* 2005;102:2454-2459.
- Yao Z, Weinberg Z, Ruzzo WL. CMfinder—a covariance model based RNA motif finding algorithm. *Bioinformatics.* 2006;22:445-452.
- Ancel LW, Fontana W. Plasticity, evolvability, and modularity in RNA. *J Exp Zool.* 2000;288:242-283.
- Pollard KS, Salama SR, King B, et al. Forces shaping the fastest evolving regions in the human genome. *PLoS Genet.* 2006;2:e168.
- Pollard KS, Salama SR, Lambert N, et al. An RNA gene expressed during cortical development evolved rapidly in humans. *Nature.* 2006;443:167.
- Walter Costa MB, Honer Zu Siederdisen 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.
- Walter CMB, Höner zu Siederdisen 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.
- 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.
- Rivas E, Eddy SR. Noncoding RNA gene detection using comparative sequence analysis. *BMC Bioinformatics.* 2001;2:8.
- 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.
- Pedersen JS, Bejerano G, Siepel A, et al. Identification and classification of conserved RNA secondary structures in the human genome. *PLoS Comput Biol.* 2006;2:e33.
- Gesell T, Washietl S. Dinucleotide controlled null models for comparative RNA gene prediction. *BMC Bioinformatics.* 2008;9:248.