

The Origin of the 5S Ribosomal RNA Molecule Could Have Been Caused by a Single Inverse Duplication: Strong Evidence from Its Sequences

Sergio Branciamore · Massimo Di Giulio

Received: 28 October 2011 / Accepted: 23 March 2012 / Published online: 11 April 2012
© Springer Science+Business Media, LLC 2012

Abstract The secondary structure of the 5S ribosomal RNA (5S rRNA) molecule shows a high degree of symmetry. In order to explain the origin of this symmetry, it has been conjectured that one half of the 5S rRNA molecule was its precursor and that an indirect duplication of this precursor created the other half and thus the current symmetry of the molecule. Here, we have subjected to an empirical test both the indirect duplication model, analysing a total of 684 5S rRNA sequences for complementarity between the two halves of the 5S rRNA, and the direct duplication model analysing in this case the similarity between the two halves of this molecule. In intra- and inter-molecule and intra- and inter-domain comparisons, we find a high statistical support to the hypothesis of a complementarity relationship between the two halves of the 5S rRNA molecule, denying vice versa the hypothesis of similarity between these halves. Therefore, these observations corroborate the indirect duplication model at the expense of the direct duplication model, as reason of the origin of the 5S rRNA molecule. More generally, we discuss and favour the hypothesis that all RNAs and proteins, which present symmetry, did so through gene duplication and not by gradualistic accumulation of few monomers or

segments of molecule into a gradualistic growth process. This would be the consequence of the very high propensity that nucleic acids have to be subjected to duplications.

Keywords Duplication models · Indirect and direct duplications · Gradualistic and punctuated evolution of molecules · Symmetrical molecules · Evolutionary precursors

Introduction and Hypothesis

In its more general meaning, the gene duplication should have been one of the fundamental mechanisms on which was played a large part of the origin and evolution of biological structures on all its levels. This statement is strongly supported by the extraordinary propensity that nucleic acids have for replication and duplication. Indeed, this point of view has been universally recognised after the work of Ohno (1970), who suggested that gene and genome duplications must have been one of the primary forces guiding the adaptive evolution of genetic systems. It is obvious that gene duplication must have been more important and more operative in the very early phases of the evolution of life, those which led to the formation of fundamental molecules. That is to say, gene duplication should also be easily traceable in the genesis of several of fundamental molecules that characterise the present-day organisms and in particular into basic molecules that are made by RNA. In actual fact, there are some instances of basic molecules that seem to have originated by means of duplication as, for example, the tRNA molecule which might have been generated by means of the assembly of two hairpin-like structures of RNA (formally a direct

S. Branciamore
Department of Physics and Astronomy, University of Florence,
Largo E. Fermi 2, 50125 Florence, Italy

M. Di Giulio (✉)
Laboratory for Molecular Evolution, Institute of Genetics and
Biophysics 'Adriano Buzzati Traverso', CNR, Via P. Castellino,
111, 80131 Naples, Italy
e-mail: digiulio@igb.cnr.it

duplication) (Di Giulio 1992; Branciamore and Di Giulio 2011).

The aptitude by which nucleic acids replicate and duplicate coupled with the existence, for instance, of symmetry conditions of a given molecule which might have appeared by means of gene duplication has not prevented that gradualistic “increasing” mechanisms were responsible for the generation of some molecules through the insertion of a few monomers in evolutionary steps. For instance, Rodin et al. (2011) have suggested that the tRNA molecule originated by means of the gradual accumulation of small sections, during six evolutionary stages, in a growth of this molecule which seems gradualistic. A gradualistic point of view has also been proposed by Sun and Caetano-Anollés (2008), although their model is different in detail from that suggested by Rodin et al. (2011). Similarly, Sun and Caetano-Anollés (2009, 2010) have reported that the 5S ribosomal RNA molecule (5S rRNA) emerged through five evolutionary stages, in which it seems that a gradualistic and non-punctuated growth of this molecule was present.

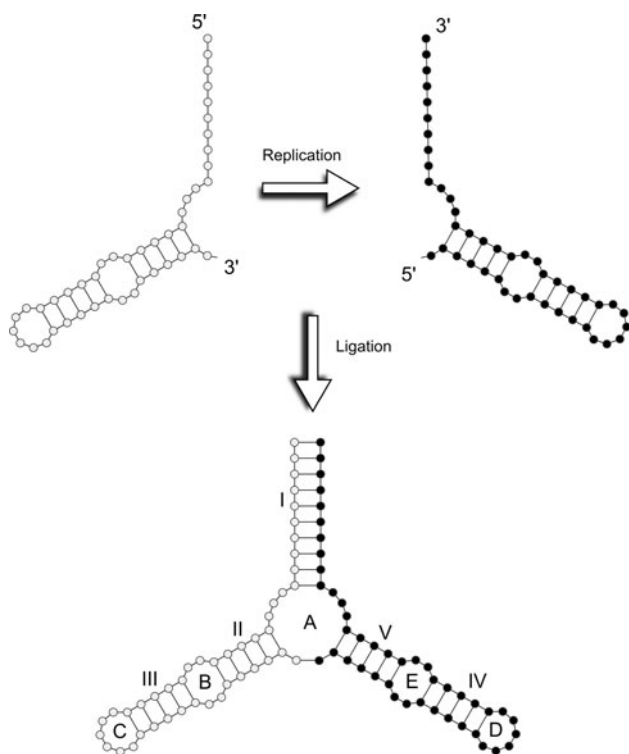


Fig. 1 A model of the origin of the 5S rRNA molecule (Di Giulio 2010). It is shown how a simple complementary copy of the 5PH of this molecule it is able to produce the 3PH, and the ligation of these two halves creates the complete secondary structure of the 5S rRNA molecule. The indicated regions of the secondary structure are consistent with the official nomenclature (see, for instances, Sun and Caetano-Anollés 2010)

However, the almost perfect symmetry of the 5S rRNA molecule (Fox and Woese 1975; Hannock and Wagner 1982; Joachimiak et al. 1990; Luehrsen and Fox 1981; Sun and Caetano-Anollés 2009; Fig. 1) seems strongly to favour the intervention of a single indirect duplication for its origin (Di Giulio 2010; Fig. 1), that is to say, its origin might have been punctuated and not gradualistic. This also seems to be the case for the origin of the tRNA molecule (Di Giulio 1992, 2004, 2006, 2009) contrarily to the suggestions of other authors (Sun and Caetano-Anollés 2008; Rodin et al. 2011). The V domain of the 23S rRNA seems also has had a segment of 110 nucleotides duplicated to place the P and A sites of the ribosome in the present evolutionary relationship (Bokov and Steinberg 2009). Furthermore, we wish to emphasise that a unique direct duplication could also be involved in the origin of the 5S rRNA molecule, depending upon the presence or the absence of the stem between the first bases at the 5' and the last at the 3'-end of this molecule, i.e. the so-called I stem (or S1) (Fig. 1), in the nascent stages of the origin of the 5S rRNA.

The apparent near-perfect symmetry of the 5S rRNA molecule (Fig. 1) is doubted by Sun and Caetano-Anollés (2010), who have provided evidence in favour of a model of gradualistic growth which would result more parsimonious than the one based on a unique inverse duplication (Di Giulio 2010). Their suggestion is particularly challenging in that it would imply that a single duplication could become less parsimonious than a gradualistic accretion for the formation of a symmetrical molecule like the 5S rRNA. This should contradict the logic that nucleic acids are highly predisposed to undergo duplications of discrete entity into the case in which there are conditions of symmetry that can be generated by means of duplication. More generally, we favour the hypothesis that a symmetrical molecule is only created by means of gene duplication. The symmetry could not arise through from gradualistic growth as the latter should be less parsimonious than the propensity of nucleic acids to undergo discrete duplications.

Thus, we are not persuaded by the suggestion of Sun and Caetano-Anollés (2010) and therefore instead have subjected the punctuated model to an empirical test, by means of its main prediction, that is to say, that of a unique indirect duplication. In other words, our model sustains the idea that the complementarity between the two halves of the 5S rRNA molecule was the cause of its origin (Di Giulio 2010; Fig. 1). We have thus analysed sequences of the 5S rRNA to see whether there are yet clues of an evolutionary relationship of complementarity between the two halves of this molecule: witnesses of the indirect duplication. At the same time, we have tested the direct duplication model which might reveal the origin of the 5S rRNA molecule, looking for, in this case, not the

complementarity but the similarity between the two halves of this molecule in the expectation of finding the two halves to be homologous.

Materials and Methods

Source of 5S rRNA Sequences

5S rRNA sequences were downloaded from Szymanski et al. (2000). The database provides information of phylogenetically conserved secondary structures of Archaeal, Bacterial and Eukaryal 5S rRNAs and multiple sequence structural alignments for each of the three domains of life. The original database was cleaned removing low quality sequences (i.e. incomplete) and we eliminated sequences which were too similar (i.e. belonging to strictly related species). Our refined database consisted of 364 sequences from Eukarya, 64 sequences from Archaea, and 256 sequences from Bacteria.

Generation of Random 5S rRNA Sequences

Each simulated 5S rRNA sequence has the following properties: (i) the same base composition and (ii) it can fold in the same secondary structure of the corresponding real 5S rRNA sequence, as reported in the database.

The algorithm used to produce a simulated sequence of a real 5S rRNA sequence consists of three steps

- (1) read the real 5S rRNA sequence from the database and evaluate the base composition. We estimate independently the base frequency for paired and unpaired regions
- (2) read the corresponding secondary structure in dot-bracket notation from the database
- (3) starting from the 5'-end read the secondary structure of the 5S rRNA at the position i th:

if position i th is a dot '.' then assign a random nucleotide N with probability proportional to the frequency of unpaired nucleotides in the real 5S rRNA.
 if position i th is open bracket '(' then find the complementary position i' th in the secondary structure, and assign one of the possible nucleotide pairs $N::N'$ to position i th and i' th with probability proportional to the frequency of paired nucleotides in the real 5S rRNA. Repeat step 3 for the whole length of the sequence (see also Branciamore and Di Giulio 2011).

Definition of the Two Halves of the 5S rRNA Molecule

According to the proposed model, the molecules of the 5S rRNA were divided into two halves: the 5' half (5PH) and

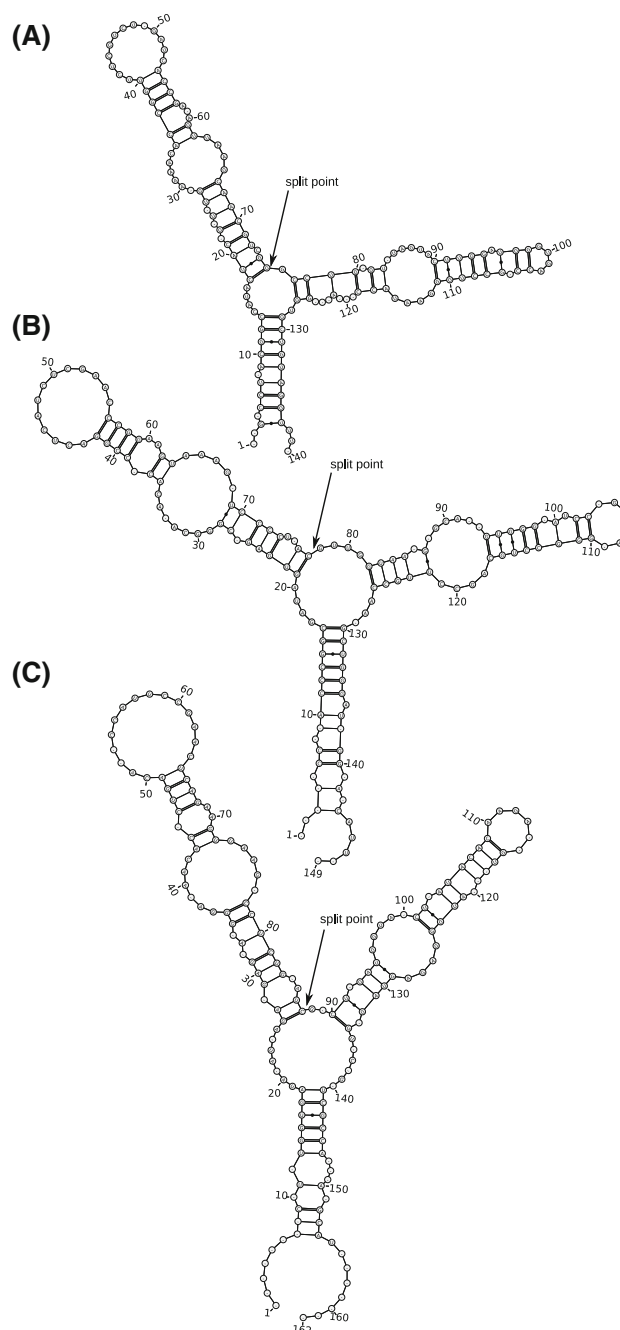


Fig. 2 The 5S rRNA consensus secondary structure for **a** Eukarya, **b** Archaea and **c** Bacteria. Arrow marks the position where the 5S rRNA molecule was divided in the 5PH and in the 3PH (3PH)

the 3' half (3PH) (Di Giulio 2010; Fig. 2). The two halves were then aligned and compared (see below).

To check the hypothesis that insertions/deletions occurred during evolution of the 5S rRNA, we removed nucleotides progressively from the 5'-end of the 5PH before proceeding to the alignment of the two halves. The

number 0, 1, 2, 3, ... indicate the number of nucleotides removed from the 5PH before the alignment.

Alignment of Two Halves of the 5S rRNA

Alignments of two halves were obtained using the software MUSCLE (Edgar 2004) implemented in a custom Perl scripts. The identity percentage calculated as the number of identical nucleotides in the alignment of the real 5S rRNA from database was compared with the one observed in a simulated sample. For testing the direct and inverse duplications, a percentage of identity between 5PH and 3PH (or 5PH and rc3PH) was calculated. A library of 1,000 simulated 5S rRNA sequences was built as described above. For each 5S rRNA sequence in the database, we calculate a $Z(I)$:

$$Z(I) = (I^{\text{real}} - I^{\text{rand}}) / s^{\text{rand}}$$

where I^{real} is the identity percentage found in the real sequence, I^{rand} is the average identity percentage found in the simulated sample and s^{rand} is the standard deviation of the percentage identity in the simulated ensemble [see Chamary and Hurst 2005 for a discussion of a variable equivalent to $Z(I)$]. The $Z(I)$ represents the number of standard deviations that the observed identity percentage in real 5S rRNA sequences differs from the one that is observed in the simulated ensemble. Positive values of the $Z(I)$ represent a similarity higher than is expected only by chance. The distribution of $Z(I)$ values should be normal with mean zero.

In the intra-molecular analysis, we compared regions of the 5S rRNA belonging to the same molecule. In the inter-molecular analysis the 5PH of a molecule was compared to the 3PH of another 5S rRNA molecule taken randomly from the remaining 5S rRNAs of the same domain of life. Finally, in the case of inter-domain analysis the two halves belong to different domains of life; for example, in the comparison between Eukaryal and Bacterial domains, the first half belongs to Eukarya whereas the second half to Bacteria.

R software package was used to perform a t test to evaluate whether the distribution of the $Z(I)$ score observed in the real 5S rRNA sample are significantly different from that observed, only by chance, in a simulated sample.

Results and Specific Discussion

We tested the hypothesis that the origin of the 5S rRNA molecule was based on duplication (direct or inverse) of an ancestral hairpin-like molecule followed by covalent bonding of the two halves to generate the actual 5S rRNA secondary structure (see Introduction and Hypothesis, and Fig. 1; Di Giulio 2010). If the hypothesis were to be correct we might find clues of the common origin of the two halves

in modern sequences of the 5S rRNA. We estimated the percentage of identity observed between the two halves in real 5S rRNA sequences and compared it to that expected only by chance in the simulated sample (Chamary and Hurst 2005). In particular, according to the consensus secondary structures reported in the 5S rRNA database, the 5S rRNA secondary structure was divided in two halves: the 5PH and the 3PH as shown in Fig. 2. The direct duplication hypothesis was tested by direct alignment of the 5PH and 3PH, whereas to test the inverse duplication hypothesis we aligned the 5PH with the reverse complementary sequence of the 3PH (rc3PH).

As the consensus secondary structures of the 5S rRNA differ in the three domains of life (Fig. 2) the analyses were done independently for Bacterial, Archaeal and Eukaryal domains. We performed an intra-molecular, inter-molecular and inter-domain comparison.

Intra-molecular Comparison

The percentage of identity between the two putative halves, in the case of the direct duplication model (5PH vs 3PH) for Archaeal, Eukaryal and Bacterial domains is not significantly higher than would be expected by chance (Tables 2, 3, 4). Considering the inverse duplication model (5PH vs rc3PH) the percentage of identity between the two halves was significant for Eukaryal and Bacterial domains (Tables 2, 4), whereas for Archaeal domain the percentage of identity was not statistically significant (Table 3) with respect to that observed in the simulated 5S rRNA sequences.

The comparison of the two putative 5PH and 3PH (or rc3PH) as such seems to support the prediction of the inverse duplication model for the origin of the 5S rRNA molecule, for Eukaryal and Bacterial domains but not for the Archaeal domain and with probabilities only marginally significant. However, before we reject the models, we observe that the length of the 5PH in modern 5S rRNAs is longer than that of the 3PH (Fig. 2; Table 1). This means that the perfect symmetry assumed by direct and inverse duplication models was lost (if ever present) during evolution of the 5S rRNA. In particular, in Archaeal and Eukaryal domains on the average 5PH is about 11 nucleotides longer than 3PH and more than 17 nucleotides in the case of the Bacterial domain (Table 1). This observation could suggest that during the evolutionary process towards the modern 5S rRNA, insertions/deletions occurred, breaking the perfect symmetry assumed by duplication models. Whether this is correct, it will be necessary to remove the 'excess' of nucleotides presented in the 5PH to correctly align with the right portions of the 3PH. More generally, the removal of nucleotides from the 5PH might favour the alignment between the 5PH and 3PH. To this end, we use

Table 1 Average lengths of the 5PH and 3PH

| | 5PH (Average number of nucleotides) | 3PH (Average number of nucleotides) |
|----------|-------------------------------------|-------------------------------------|
| Eukarya | 65.3 | 54.2 |
| Archaea | 67.6 | 56.5 |
| Bacteria | 67.1 | 49.6 |

three distinct strategies to remove nucleotides from the 5PH:

- (i) 5' removal (5PR): nucleotides from the 5PH at the 5'-end were progressively removed before the alignment with the 3PH or rc3PH. Then the percentage of identity and $Z(I)$ between the two halves was evaluated.
- (ii) 3' removal (3PR): nucleotides from the 5PH at the 3'-end were progressively removed before the alignment with the 3PH or rc3PH. Then the percentage of identity and $Z(I)$ between the two halves was evaluated.
- (iii) Random removal (RR): an increasing number of nucleotides from the 5PH were randomly selected and removed before the alignment with the 3PH or rc3PH. Then the percentage of identity and $Z(I)$ between the two halves was evaluated.

The $Z(I)$ value between the two halves is reported as a function of the number of nucleotides removed (using the different strategies described above) from the 5PH, before of its alignment with the 3PH (or rc3PH) (Fig. 3). Results for the intra-molecular comparison are showed in Fig. 3.

Initially, we examined the case of the direct duplication model. For all three domains of life, the percentage of identity between the two halves did not increase using 5PR strategy, and the percentage of identity was not higher than that observed in a sample of simulated 5S rRNA sequences (Tables 2, 3, 4). The results were similar also using the 3PR or RR strategies (data not shown). The only observed exception was for Eucaryal and Archaeal domains showing a peak with a significantly higher percentage of identity after removal of about 22 nucleotides (data not shown).

The inverse duplication model showed a different case. Indeed, under the assumption of this model, one half has been the precursor of the entire molecule, and an inverse duplication of this precursor created the complete 5S rRNA molecule (Di Giulio 2010; Fig. 1). In this case, we should expect complementarity between the 5PH and 3PH (similarity between the 5PH and rc3PH). We first describe the results using the 5PR strategy.

The Eukaryal domain (Fig. 3a; Table 2) shows a clear increase of the percentage of identity between the 5PH and rc3PH after removal of nucleotides at the 5'-end of the

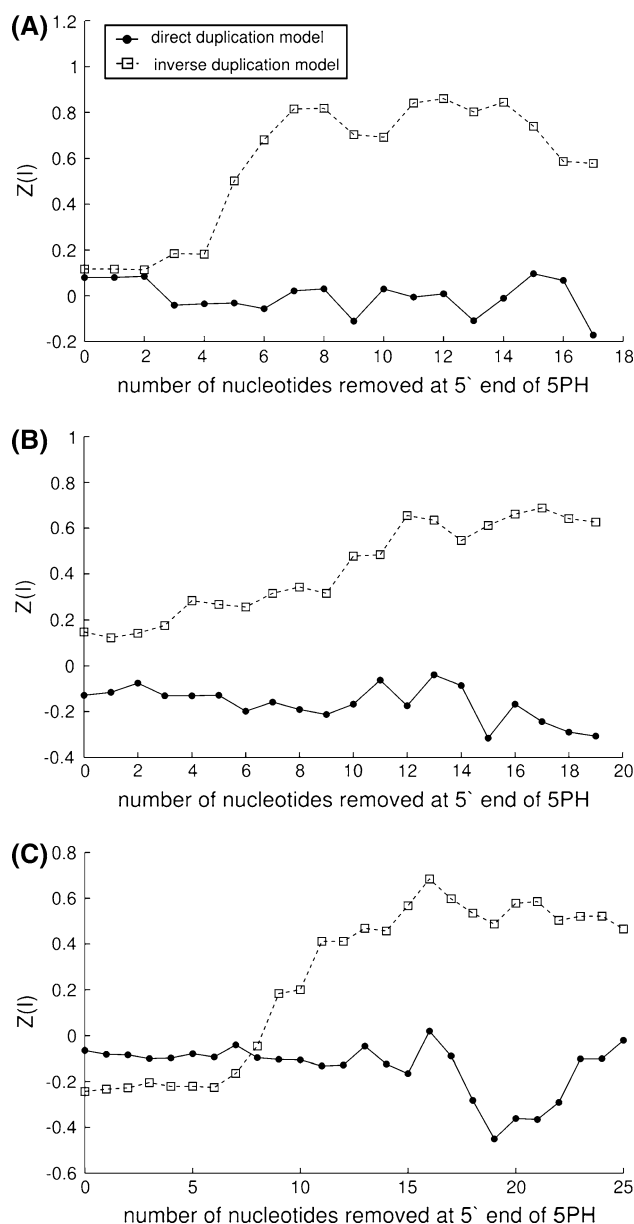


Fig. 3 Intra-molecular analysis. Variation of the $Z(I)$ score (see 'Materials and methods') as a function of nucleotides removed at the 5'-end of the 5PH (5PR strategy) in **a** Eukarya, **b** Archaea and **c** Bacteria. The solid line with filled circles represents the $Z(I)$ score observed for the direct duplication model (5PH vs 5PH). The dashed line with empty squares represents the $Z(I)$ score observed for the inverse duplication model (5PH vs rc3PH)

5PH. In particular, after removal of the first four nucleotides it is observed a clear increase in the $Z(I)$ value which reaches a maximum value of 0.82 ($P = 8.2 \times 10^{-43}$) after removal of 7–8 nucleotides, remaining almost constant up to the removal of 30 nucleotides and then starts to decrease (see Table 2 for a full list of results after removal up to 40 nucleotides).

The Archaeal domain presents a similar behaviour (Fig. 3b; Table 3). The $Z(I)$ value grows as nucleotides

Table 2 Eukaryal intra-molecular analysis utilising the 5PR strategy

| Nucleotides removed at the 5'-end of 5PH | Direct duplication model | | Inverse duplication model | |
|--|--------------------------|------------------------|---------------------------|------------------------|
| | $Z(I)$ | P^* | $Z(I)$ | P^* |
| 0 | 0.08 | 0.13 | 0.12 | 0.005 |
| 1 | 0.08 | 0.13 | 0.12 | 0.005 |
| 2 | 0.08 | 0.10 | 0.11 | 0.007 |
| 3 | -0.04 | 0.41 | 0.18 | 1.6×10^{-4} |
| 4 | -0.03 | 0.47 | 0.18 | 1.8×10^{-4} |
| 5 | -0.03 | 0.52 | 0.5 | 4.64×10^{-22} |
| 6 | -0.06 | 0.26 | 0.68 | 9.40×10^{-35} |
| 7 | 0.02 | 0.67 | 0.82 | 8.19×10^{-43} |
| 8 | 0.03 | 0.55 | 0.82 | 6.73×10^{-43} |
| 9 | -0.11 | 0.02 | 0.7 | 2.90×10^{-33} |
| 10 | 0.03 | 0.52 | 0.69 | 6.82×10^{-31} |
| 11 | -0.02 | 0.90 | 0.84 | 2.20×10^{-41} |
| 12 | 0.02 | 0.85 | 0.86 | 1.25×10^{-37} |
| 13 | -0.11 | 0.04 | 0.8 | 4.46×10^{-38} |
| 14 | -0.01 | 0.83 | 0.84 | 7.18×10^{-40} |
| 15 | 0.01 | 0.04 | 0.74 | 2.48×10^{-35} |
| 16 | 0.07 | 0.20 | 0.59 | 1.07×10^{-23} |
| 17 | -0.17 | 6.80×10^{-4} | 0.58 | 2.92×10^{-26} |
| 18 | -0.13 | 0.01 | 0.79 | 1.95×10^{-44} |
| 19 | -0.13 | 3.36×10^{-3} | 0.94 | 6.83×10^{-58} |
| 20 | -0.17 | 2.20×10^{-4} | 0.85 | 1.80×10^{-49} |
| 21 | -0.18 | 6.04×10^{-5} | 0.87 | 1.09×10^{-52} |
| 22 | -0.18 | 7.46×10^{-5} | 0.88 | 7.57×10^{-54} |
| 23 | -0.36 | 4.71×10^{-15} | 0.87 | 1.07×10^{-56} |
| 24 | -0.36 | 4.09×10^{-15} | 0.87 | 1.04×10^{-56} |
| 25 | -0.5 | 1.74×10^{-24} | 0.8 | 1.72×10^{-45} |
| 26 | -0.5 | 1.74×10^{-24} | 0.8 | 1.87×10^{-45} |
| 27 | -0.42 | 8.01×10^{-21} | 0.79 | 1.51×10^{-46} |
| 28 | -0.63 | 1.40×10^{-39} | 0.91 | 7.32×10^{-56} |
| 29 | -0.63 | 1.23×10^{-39} | 0.91 | 7.18×10^{-56} |
| 30 | -0.53 | 4.77×10^{-28} | 0.83 | 2.55×10^{-47} |
| 31 | -0.56 | 4.64×10^{-31} | 0.64 | 2.94×10^{-31} |
| 32 | -0.34 | 3.95×10^{-13} | 0.63 | 5.21×10^{-35} |
| 33 | -0.45 | 3.93×10^{-20} | 0.43 | 2.79×10^{-19} |
| 34 | -0.22 | 5.33×10^{-6} | 0.49 | 1.33×10^{-25} |
| 35 | -0.33 | 4.68×10^{-12} | 0.39 | 2.28×10^{-17} |
| 36 | -0.37 | 7.11×10^{-14} | 0.35 | 6.57×10^{-15} |
| 37 | -0.34 | 1.84×10^{-13} | 0.26 | 2.43×10^{-8} |
| 38 | -0.37 | 6.37×10^{-15} | 0.18 | 8.59×10^{-5} |
| 39 | -0.52 | 4.54×10^{-25} | 0.2 | 1.65×10^{-6} |
| 40 | -0.42 | 7.90×10^{-17} | -0.02 | 0.7 |

are removed from the 5'-end of the 5PH and reaches a maximum value after removal of 18–19 nucleotides ($Z(I) = 0.63$, $P = 1.39 \times 10^{-9}$) (Table 3).

Similar to that previously observed in Eukaryal and Archaeal domains also the Bacterial domain shows the

$Z(I)$ value which increases after removal of nucleotides at the 5'-end of the 5PH (Fig. 3c; Table 4). The comparison of the 5PH versus rc3PH in the Bacterial domain shows $Z(I)$ values < 0 , but after removal of 11 nucleotides the $Z(I)$ value suddenly increases, and reaches a maximum

Table 3 Archaeal intra-molecular analysis utilising the 5PR strategy

| Nucleotides removed at the 5'-end of the 5PH | Direct duplication model | | Inverse duplication model | |
|--|--------------------------|-----------------------|---------------------------|-----------------------|
| | $Z(I)$ | P^* | $Z(I)$ | P^* |
| 0 | -0.13 | 0.27 | 0.15 | 0.16 |
| 1 | -0.12 | 0.33 | 0.12 | 0.23 |
| 2 | -0.07 | 0.51 | 0.14 | 0.17 |
| 3 | -0.13 | 0.28 | 0.18 | 0.11 |
| 4 | -0.13 | 0.25 | 0.28 | 0.01 |
| 5 | -0.13 | 0.27 | 0.27 | 0.01 |
| 6 | -0.2 | 0.06 | 0.26 | 0.04 |
| 7 | -0.16 | 0.19 | 0.32 | 0.01 |
| 8 | -0.19 | 0.11 | 0.34 | 3.55×10^{-3} |
| 9 | -0.21 | 0.06 | 0.32 | 0.01 |
| 10 | -0.17 | 0.11 | 0.48 | 4.97×10^{-5} |
| 11 | -0.06 | 0.53 | 0.48 | 2.65×10^{-5} |
| 12 | -0.17 | 0.07 | 0.66 | 3.53×10^{-7} |
| 13 | -0.04 | 0.70 | 0.63 | 3.11×10^{-7} |
| 14 | -0.09 | 0.41 | 0.55 | 9.29×10^{-5} |
| 15 | -0.31 | 0.00 | 0.61 | 1.95×10^{-7} |
| 16 | -0.17 | 0.14 | 0.66 | 5.58×10^{-7} |
| 17 | -0.24 | 0.02 | 0.69 | 2.65×10^{-8} |
| 18 | -0.29 | 0.01 | 0.64 | 3.26×10^{-9} |
| 19 | -0.31 | 1.5×10^{-3} | 0.63 | 1.39×10^{-9} |
| 20 | -0.14 | 0.14 | 0.5 | 1.27×10^{-4} |
| 21 | -0.44 | 9.39×10^{-5} | 0.55 | 1.16×10^{-4} |
| 22 | -0.3 | 0.01 | 0.4 | 9.22×10^{-4} |
| 23 | -0.19 | 0.05 | 0.42 | 7.34×10^{-4} |
| 24 | -0.39 | 3.15×10^{-4} | 0.25 | 0.05 |
| 25 | -0.35 | 6.79×10^{-4} | 0.4 | 3.11×10^{-3} |
| 26 | -0.35 | 8.53×10^{-4} | 0.41 | 1.95×10^{-3} |
| 27 | -0.12 | 0.23 | 0.29 | 0.03 |
| 28 | -0.07 | 0.58 | 0.26 | 0.06 |
| 29 | 0.08 | 0.55 | 0.04 | 0.77 |
| 30 | -0.22 | 0.07 | -0.14 | 0.24 |
| 31 | -0.35 | 4.90×10^{-3} | -0.13 | 0.27 |
| 32 | -0.18 | 0.13 | -0.26 | 0.03 |
| 33 | -0.13 | 0.24 | -0.28 | 0.02 |
| 34 | -0.14 | 0.22 | -0.24 | 0.04 |
| 35 | -0.08 | 0.4 | -0.24 | 0.03 |
| 36 | -0.03 | 0.8 | -0.21 | 0.02 |
| 37 | -0.02 | 0.88 | -0.11 | 0.31 |
| 38 | -0.02 | 0.86 | -0.11 | 0.33 |
| 39 | 0 | 0.96 | 0.05 | 0.61 |
| 40 | 0.39 | 1.68×10^{-3} | -0.02 | 0.88 |

value of $Z(I) = 0.68$ ($P = 8.76 \times 10^{-14}$) when 16 nucleotides were deleted (Table 4).

The situation is different if nucleotides were removed starting from the 3'-end of the 5PH (3PR strategy). In this case, the alignment of the 5PH and rc3PH did

not show the similarity expected for the inverse duplication model (data not shown, see also Table 11A). The RR of nucleotides from the 5PH by means of the RR strategy gives results equivalent to that observed using the 5PR strategy (data not shown, see also Table 11).

Table 4 Bacterial intra-molecular analysis utilising the 5PR strategy

| Nucleotides removed at the 5'-end of the 5PH | Direct duplication model | | Inverse duplication model | |
|--|--------------------------|------------------------|---------------------------|------------------------|
| | Z(I) | P* | Z(I) | P* |
| 0 | -0.06 | 0.41 | -0.24 | 1.48×10^{-3} |
| 1 | -0.08 | 0.3 | -0.23 | 2.18×10^{-3} |
| 2 | -0.08 | 0.29 | -0.23 | 2.86×10^{-3} |
| 3 | -0.1 | 0.21 | -0.2 | 0.01 |
| 4 | -0.1 | 0.22 | -0.22 | 3.92×10^{-3} |
| 5 | -0.08 | 0.32 | -0.22 | 4.21×10^{-3} |
| 6 | -0.09 | 0.25 | -0.23 | 3.44×10^{-3} |
| 7 | -0.04 | 0.62 | -0.16 | 0.03 |
| 8 | -0.1 | 0.24 | -0.05 | 0.56 |
| 9 | -0.1 | 0.2 | 0.18 | 0.02 |
| 10 | -0.11 | 0.19 | 0.2 | 0.01 |
| 11 | -0.13 | 0.11 | 0.41 | 1.05×10^{-6} |
| 12 | -0.13 | 0.12 | 0.41 | 1.10×10^{-6} |
| 13 | -0.05 | 0.59 | 0.47 | 3.78×10^{-7} |
| 14 | -0.12 | 0.16 | 0.46 | 4.92×10^{-7} |
| 15 | -0.17 | 0.05 | 0.57 | 5.46×10^{-10} |
| 16 | 0.02 | 0.8 | 0.68 | 8.76×10^{-14} |
| 17 | -0.09 | 0.27 | 0.6 | 7.71×10^{-11} |
| 18 | -0.28 | 5.69×10^{-9} | 0.53 | 4.32×10^{-11} |
| 19 | -0.45 | 1.92×10^{-9} | 0.49 | 1.06×10^{-8} |
| 20 | -0.36 | 4.87×10^{-7} | 0.58 | 2.32×10^{-13} |
| 21 | -0.37 | 3.32×10^{-7} | 0.59 | 9.18×10^{-14} |
| 22 | -0.29 | 2.68×10^{-5} | 0.5 | 2.14×10^{-10} |
| 23 | -0.1 | 0.15 | 0.52 | 1.91×10^{-10} |
| 24 | -0.1 | 0.15 | 0.52 | 1.82×10^{-10} |
| 25 | -0.02 | 0.76 | 0.47 | 4.65×10^{-9} |
| 26 | -0.28 | 2.31×10^{-5} | 0.37 | 2.35×10^{-6} |
| 27 | -0.3 | 5.08×10^{-5} | 0.38 | 4.14×10^{-7} |
| 28 | -0.3 | 5.08×10^{-5} | 0.38 | 4.14×10^{-7} |
| 29 | -0.48 | 1.31×10^{-9} | 0.35 | 8.65×10^{-6} |
| 30 | -0.48 | 2.23×10^{-11} | 0.36 | 3.40×10^{-6} |
| 31 | -0.42 | 7.24×10^{-9} | 0.26 | 7.26×10^{-4} |
| 32 | -0.42 | 7.24×10^{-9} | 0.26 | 7.26×10^{-4} |
| 33 | -0.5 | 5.53×10^{-12} | 0.29 | 1.15×10^{-4} |
| 34 | -0.5 | 5.29×10^{-12} | 0.28 | 1.35×10^{-4} |
| 35 | -0.39 | 5.15×10^{-8} | 0.27 | 3.69×10^{-4} |
| 36 | -0.42 | 1.05×10^{-8} | 0.36 | 2.98×10^{-6} |
| 37 | -0.51 | 1.17×10^{-11} | 0.37 | 5.12×10^{-6} |
| 38 | -0.46 | 5.71×10^{-10} | 0.32 | 2.62×10^{-5} |
| 39 | -0.45 | 9.18×10^{-10} | 0.32 | 2.62×10^{-5} |
| 40 | -0.3 | 1.48×10^{-5} | 0.38 | 5.44×10^{-7} |

Inter-molecular Comparison

The inter-molecular comparison should reduce correlation biases eventually present in the intra-molecular analysis due to interactions between different regions of the same molecule. We repeated the analysis described in the

previous session. However, in this case the halves compared (the 5PH and 3PH) do not belong to the same molecule but were chosen from two different 5S rRNAs.

Results obtained for the inter-molecular comparison overlap with that of the intra-molecular analysis. Identity percentages of the two aligned halves do not support direct

Table 5 Eukaryal inter-molecular analysis utilising the 5PR strategy

| Nucleotides removed at the 5'-end of the 5PH | Direct duplication model | | Inverse duplication model | |
|--|--------------------------|-----------------------|---------------------------|-----------------------|
| | Z(I) | P* | Z(I) | P* |
| 0 | 0.016 | 0.738 | 0.529 | 2.3×10^{-26} |
| 1 | 0.016 | 0.735 | 0.530 | 2.1×10^{-26} |
| 2 | 0.020 | 0.685 | 0.531 | 1.8×10^{-26} |
| 3 | -0.097 | 0.032 | 0.476 | 5.3×10^{-25} |
| 4 | -0.094 | 0.038 | 0.469 | 1.3×10^{-24} |
| 5 | -0.039 | 0.427 | 0.509 | 7.6×10^{-26} |
| 6 | -0.117 | 0.017 | 0.591 | 1.7×10^{-32} |
| 7 | -0.037 | 0.423 | 0.644 | 6.2×10^{-36} |
| 8 | -0.020 | 0.657 | 0.643 | 7.6×10^{-36} |
| 9 | -0.157 | 0.001 | 0.508 | 1.7×10^{-23} |
| 10 | -0.122 | 0.010 | 0.582 | 1.4×10^{-27} |
| 11 | -0.085 | 0.062 | 0.639 | 1.5×10^{-29} |
| 12 | -0.088 | 0.044 | 0.581 | 2.9×10^{-25} |
| 13 | -0.217 | 1.9×10^{-5} | 0.694 | 3.1×10^{-35} |
| 14 | -0.158 | 0.001 | 0.679 | 1.8×10^{-35} |
| 15 | -0.010 | 0.825 | 0.553 | 2.1×10^{-25} |
| 16 | -0.075 | 0.116 | 0.501 | 1.1×10^{-22} |
| 17 | -0.251 | 3.5×10^{-7} | 0.444 | 2.4×10^{-18} |
| 18 | -0.268 | 7.3×10^{-9} | 0.621 | 1.7×10^{-31} |
| 19 | -0.208 | 1.1×10^{-6} | 0.754 | 1.6×10^{-42} |
| 20 | -0.162 | 4.2×10^{-4} | 0.710 | 1.2×10^{-39} |
| 21 | -0.142 | 9.8×10^{-4} | 0.745 | 4.3×10^{-44} |
| 22 | -0.140 | 0.001 | 0.746 | 2.2×10^{-44} |
| 23 | -0.275 | 3.0×10^{-9} | 0.681 | 3.0×10^{-37} |
| 24 | -0.279 | 1.6×10^{-9} | 0.679 | 6.6×10^{-37} |
| 25 | -0.411 | 2.9×10^{-17} | 0.614 | 1.3×10^{-32} |
| 26 | -0.411 | 2.9×10^{-17} | 0.613 | 1.6×10^{-32} |
| 27 | -0.387 | 9.1×10^{-20} | 0.604 | 3.2×10^{-32} |
| 28 | -0.548 | 1.6×10^{-30} | 0.645 | 5.1×10^{-35} |
| 29 | -0.552 | 6.5×10^{-31} | 0.644 | 7.1×10^{-35} |
| 30 | -0.499 | 1.8×10^{-29} | 0.530 | 2.4×10^{-23} |
| 31 | -0.472 | 2.6×10^{-25} | 0.445 | 4.8×10^{-19} |
| 32 | -0.369 | 7.2×10^{-17} | 0.359 | 9.0×10^{-13} |
| 33 | -0.458 | 3.8×10^{-22} | 0.228 | 1.7×10^{-6} |
| 34 | -0.264 | 5.2×10^{-8} | 0.181 | 1.6×10^{-4} |
| 35 | -0.276 | 1.1×10^{-8} | 0.050 | 0.267 |
| 36 | -0.362 | 4.9×10^{-13} | -0.036 | 0.433 |
| 37 | -0.280 | 2.6×10^{-9} | -0.071 | 0.128 |
| 38 | -0.356 | 1.4×10^{-12} | -0.076 | 0.120 |
| 39 | -0.435 | 6.4×10^{-18} | -0.112 | 0.018 |
| 40 | -0.303 | 2.3×10^{-9} | -0.293 | 1.4×10^{-9} |

duplication model (for the 5PR, 3PR and RR strategies) (Tables 5, 6, 7). As in the previous case exceptions were observed for Eucaryal and Archeal domains showing a peak with a significantly higher percentage of identity after the removal of about 22 nucleotides, using the 3PR

strategy. However, the removal of nucleotides at the 5'-end of the 5PH (using the 5PR, 3PR and RR strategies), before alignment with the rc3PH, clearly support the inverse duplication model (see Tables 5, 6, 7; Fig. 4) (data not shown for the 3PR and RR strategies, see Table 11).

Table 6 Archaeal inter-molecular analysis utilising the 5PR strategy

| Nucleotides removed at the 5'-end of the 5PH | Direct duplication model | | Inverse duplication model | |
|--|--------------------------|----------------------|---------------------------|-----------------------|
| | Z(I) | P* | Z(I) | P* |
| 0 | -0.387 | 0.003 | 0.750 | 1.5×10^{-8} |
| 1 | -0.390 | 0.003 | 0.770 | 2.5×10^{-9} |
| 2 | -0.415 | 0.002 | 0.739 | 1.6×10^{-8} |
| 3 | -0.292 | 0.023 | 0.703 | 2.1×10^{-8} |
| 4 | -0.330 | 0.007 | 0.679 | 7.1×10^{-8} |
| 5 | -0.334 | 0.008 | 0.704 | 1.9×10^{-8} |
| 6 | -0.369 | 0.004 | 0.718 | 8.0×10^{-9} |
| 7 | -0.346 | 0.004 | 0.730 | 3.2×10^{-8} |
| 8 | -0.334 | 0.004 | 0.722 | 3.6×10^{-8} |
| 9 | -0.349 | 0.003 | 0.717 | 4.1×10^{-8} |
| 10 | -0.441 | 6.0×10^{-5} | 0.810 | 4.1×10^{-10} |
| 11 | -0.331 | 0.002 | 0.823 | 4.7×10^{-9} |
| 12 | -0.204 | 0.038 | 0.793 | 9.8×10^{-9} |
| 13 | -0.190 | 0.053 | 0.762 | 6.7×10^{-8} |
| 14 | -0.092 | 0.362 | 0.835 | 5.1×10^{-8} |
| 15 | -0.302 | 0.003 | 0.721 | 2.9×10^{-7} |
| 16 | -0.312 | 0.005 | 0.555 | 1.8×10^{-5} |
| 17 | -0.197 | 0.072 | 0.590 | 3.9×10^{-6} |
| 18 | -0.249 | 0.022 | 0.634 | 4.3×10^{-8} |
| 19 | -0.288 | 0.009 | 0.599 | 5.0×10^{-8} |
| 20 | -0.010 | 0.927 | 0.462 | 4.1×10^{-4} |
| 21 | -0.347 | 2.9×10^{-4} | 0.326 | 0.010 |
| 22 | -0.200 | 0.040 | 0.352 | 0.003 |
| 23 | -0.190 | 0.027 | 0.273 | 0.034 |
| 24 | -0.246 | 0.010 | 0.242 | 0.061 |
| 25 | -0.204 | 0.053 | 0.278 | 0.031 |
| 26 | -0.211 | 0.047 | 0.278 | 0.031 |
| 27 | -0.140 | 0.192 | 0.321 | 0.014 |
| 28 | -0.023 | 0.831 | 0.356 | 0.007 |
| 29 | 0.044 | 0.713 | 0.319 | 0.012 |
| 30 | -0.403 | 0.002 | 0.072 | 0.586 |
| 31 | -0.274 | 0.024 | 0.136 | 0.239 |
| 32 | -0.161 | 0.189 | 0.020 | 0.872 |
| 33 | -0.179 | 0.137 | -5.7×10^{-4} | 0.996 |
| 34 | -0.222 | 0.060 | -0.009 | 0.943 |
| 35 | -0.244 | 0.036 | 0.133 | 0.282 |
| 36 | -0.194 | 0.096 | 0.050 | 0.680 |
| 37 | -0.206 | 0.051 | 0.222 | 0.067 |
| 38 | -0.197 | 0.064 | 0.224 | 0.065 |
| 39 | -0.086 | 0.409 | 0.223 | 0.050 |
| 40 | 0.313 | 0.009 | 0.364 | 0.003 |

Inter-domain Comparison

To further substantiate these results, we also compared the two halves of the 5S rRNA molecule choosing them from different domains of life. The finding of traces of

a common origin between the two halves should represent a severe test for the proposed models, given the divergence in the 5S rRNA secondary structures between Eukaryal, Archaeal and Bacterial domains (Fig. 2).

Table 7 Bacterial inter-molecular analysis utilising the 5PR strategy

| Nucleotides removed at the 5'-end of the 5PH | Direct duplication model | | Inverse duplication model | |
|--|--------------------------|-----------------------|---------------------------|-----------------------|
| | Z(I) | P* | Z(I) | P* |
| 0 | -0.023 | 0.773 | 0.272 | 5.3×10^{-4} |
| 1 | -0.028 | 0.730 | 0.272 | 5.0×10^{-4} |
| 2 | -0.035 | 0.666 | 0.274 | 4.7×10^{-4} |
| 3 | -0.032 | 0.686 | 0.267 | 6.9×10^{-4} |
| 4 | -0.033 | 0.683 | 0.268 | 6.4×10^{-4} |
| 5 | -0.027 | 0.734 | 0.277 | 4.1×10^{-4} |
| 6 | -0.018 | 0.825 | 0.257 | 9.3×10^{-4} |
| 7 | 0.021 | 0.803 | 0.209 | 0.009 |
| 8 | -0.052 | 0.533 | 0.225 | 0.004 |
| 9 | 0.008 | 0.919 | 0.271 | 7.5×10^{-4} |
| 10 | 8.0×10^{-4} | 0.992 | 0.276 | 5.5×10^{-4} |
| 11 | 0.017 | 0.833 | 0.251 | 9.8×10^{-4} |
| 12 | 0.019 | 0.816 | 0.255 | 7.7×10^{-4} |
| 13 | 0.173 | 0.025 | 0.371 | 1.3×10^{-5} |
| 14 | 0.130 | 0.094 | 0.416 | 1.1×10^{-6} |
| 15 | 0.055 | 0.487 | 0.406 | 1.9×10^{-6} |
| 16 | 0.174 | 0.023 | 0.506 | 3.6×10^{-9} |
| 17 | 0.126 | 0.102 | 0.553 | 1.2×10^{-9} |
| 18 | -0.029 | 0.695 | 0.497 | 8.6×10^{-9} |
| 19 | -0.157 | 0.016 | 0.550 | 2.3×10^{-10} |
| 20 | -0.018 | 0.796 | 0.611 | 7.9×10^{-12} |
| 21 | -0.020 | 0.771 | 0.612 | 7.2×10^{-12} |
| 22 | -0.005 | 0.946 | 0.546 | 2.7×10^{-11} |
| 23 | 0.056 | 0.427 | 0.567 | 4.4×10^{-12} |
| 24 | 0.055 | 0.436 | 0.561 | 5.0×10^{-12} |
| 25 | 0.235 | 0.004 | 0.484 | 5.4×10^{-11} |
| 26 | -0.066 | 0.377 | 0.379 | 3.0×10^{-7} |
| 27 | 0.016 | 0.841 | 0.357 | 7.2×10^{-7} |
| 28 | 0.016 | 0.841 | 0.357 | 7.2×10^{-7} |
| 29 | -0.239 | 0.003 | 0.415 | 7.0×10^{-9} |
| 30 | -0.340 | 3.4×10^{-6} | 0.415 | 8.7×10^{-9} |
| 31 | -0.262 | 8.1×10^{-4} | 0.288 | 1.2×10^{-4} |
| 32 | -0.262 | 8.1×10^{-4} | 0.288 | 1.2×10^{-4} |
| 33 | -0.391 | 4.5×10^{-7} | 0.308 | 8.7×10^{-6} |
| 34 | -0.397 | 3.1×10^{-7} | 0.309 | 8.0×10^{-6} |
| 35 | -0.282 | 1.3×10^{-4} | 0.343 | 1.3×10^{-5} |
| 36 | -0.381 | 3.5×10^{-7} | 0.433 | 3.8×10^{-8} |
| 37 | -0.497 | 1.6×10^{-11} | 0.413 | 1.5×10^{-7} |
| 38 | -0.416 | 1.2×10^{-7} | 0.470 | 1.5×10^{-8} |
| 39 | -0.414 | 1.4×10^{-7} | 0.470 | 1.4×10^{-8} |
| 40 | -0.271 | 1.2×10^{-4} | 0.404 | 1.2×10^{-6} |

In this inter-domain comparison the 5PH of Eukaryal domain was compared to the 3PH of Archaeal domain (Fig. 5a; Table 8) or to that of Bacterial domain (Fig. 5b; Table 9), and the 5PH of the Bacterial domain was compared to the 3PH of the Archaeal domain (Fig. 5c;

Table 10). The 3PR and RR strategies (data not shown, see Table 11) resemble that previously described for intra- and inter-molecular comparisons. Here, we point out 5PR strategy comparisons for different domains of life.

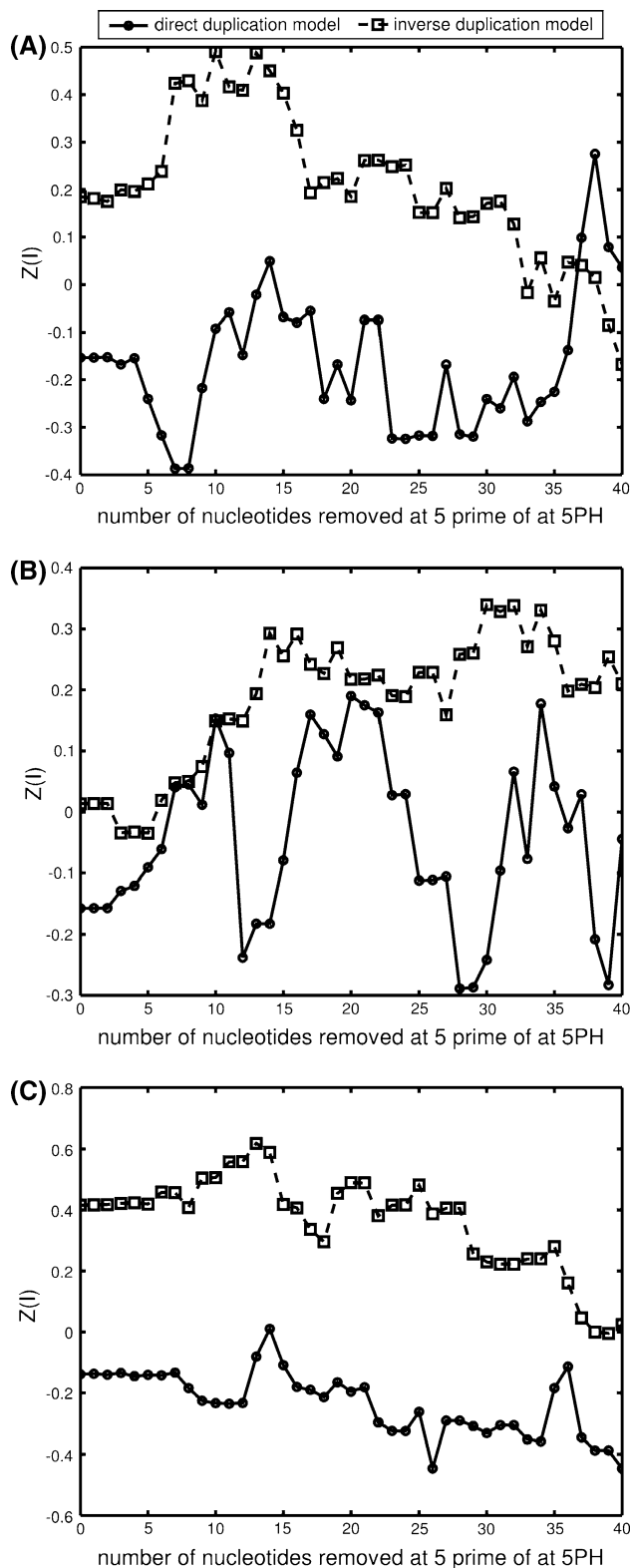
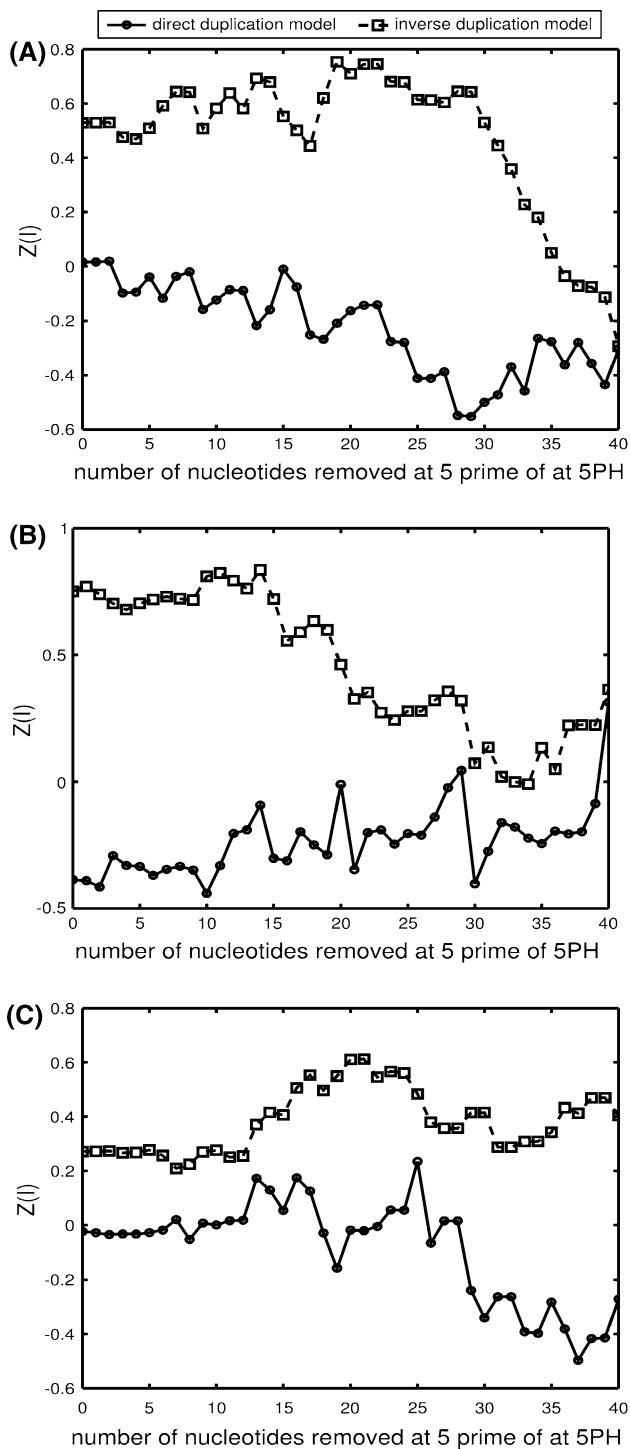


Fig. 4 Inter-molecular analysis. Variation of the $Z(I)$ score (see ‘Materials and methods’) as a function of nucleotides removed at the 5′-end of the 5PH (5PR strategy) in **a** Eukarya, **b** Archaea and **c** Bacteria. The *solid line with filled circles* represents the $Z(I)$ score observed for the direct duplication model (5PH vs 5PH). The *dashed line with empty squares* represents the $Z(I)$ score observed for the inverse duplication model (5PH vs rc3PH)

Eukaryal and Archaeal domains show that: (i) the direct duplication model is not supported by percentage of identity analysis (Table 8), whereas (ii) the inverse duplication model receives a significant support (Table 8). 5S rRNA

Fig. 5 Inter-domain analysis. Variation of the $Z(I)$ score (see ‘Materials and methods’) as a function of nucleotides removed at the 5'-end of the 5PH (5PR strategy) in **a** 5PH of Eukarya versus 3PH of Archaea, **b** 5PH of Eukarya versus 3PH of Bacteria, **c** 5PH of Bacteria versus 3PH of Archaea. The *solid line with filled circles* represents the $Z(I)$ score observed for the direct duplication model (5PH vs 3PH). The *dashed line with empty squares* represents the $Z(I)$ score observed for the inverse duplication model (5PH vs rc3PH)

molecules in Eukaryal and Bacterial domains present the highest structural divergence (Fig. 2). Results obtained substantially confirm what already observed, although the increase in the percentage of identity between the two halves is less evident, but still statistically significant, after the removal of 13 nucleotides (Table 9).

Table 8 Eukaryal versus Archaeal inter-domain analysis utilising the 5PR strategy

| Nucleotides removed at the 5'-end of the 5PH | Direct Duplication Model | | Inverse Duplication Model | |
|--|--------------------------|-----------------------|---------------------------|-----------------------|
| | $Z(I)$ | P^* | $Z(I)$ | P^* |
| 0 | -0.153 | 0.003 | 0.185 | 1.2×10^{-4} |
| 1 | -0.153 | 0.003 | 0.181 | 1.6×10^{-4} |
| 2 | -0.152 | 0.003 | 0.175 | 2.8×10^{-4} |
| 3 | -0.167 | 0.001 | 0.200 | 1.7×10^{-4} |
| 4 | -0.154 | 0.003 | 0.196 | 2.1×10^{-4} |
| 5 | -0.240 | 1.8×10^{-6} | 0.212 | 1.7×10^{-4} |
| 6 | -0.317 | 1.5×10^{-10} | 0.238 | 3.1×10^{-5} |
| 7 | -0.387 | 5.6×10^{-13} | 0.424 | 5.9×10^{-13} |
| 8 | -0.386 | 5.9×10^{-13} | 0.430 | 3.1×10^{-13} |
| 9 | -0.217 | 1.1×10^{-5} | 0.388 | 8.6×10^{-12} |
| 10 | -0.092 | 0.061 | 0.491 | 1.4×10^{-18} |
| 11 | -0.058 | 0.251 | 0.417 | 1.7×10^{-14} |
| 12 | -0.148 | 0.004 | 0.409 | 4.4×10^{-15} |
| 13 | -0.021 | 0.665 | 0.488 | 2.2×10^{-19} |
| 14 | 0.049 | 0.340 | 0.451 | 3.4×10^{-17} |
| 15 | -0.068 | 0.196 | 0.403 | 3.0×10^{-13} |
| 16 | -0.080 | 0.106 | 0.325 | 1.7×10^{-8} |
| 17 | -0.055 | 0.231 | 0.193 | 8.6×10^{-4} |
| 18 | -0.240 | 1.6×10^{-6} | 0.215 | 1.3×10^{-4} |
| 19 | -0.168 | 4.2×10^{-4} | 0.224 | 3.3×10^{-5} |
| 20 | -0.243 | 1.8×10^{-6} | 0.186 | 7.5×10^{-4} |
| 21 | -0.074 | 0.110 | 0.262 | 1.3×10^{-6} |
| 22 | -0.074 | 0.111 | 0.262 | 1.3×10^{-6} |
| 23 | -0.323 | 1.3×10^{-11} | 0.249 | 4.7×10^{-6} |
| 24 | -0.324 | 1.2×10^{-11} | 0.252 | 4.0×10^{-6} |
| 25 | -0.317 | 1.3×10^{-10} | 0.152 | 0.003 |
| 26 | -0.318 | 1.2×10^{-10} | 0.151 | 0.003 |
| 27 | -0.168 | 3.8×10^{-4} | 0.202 | 3.6×10^{-5} |
| 28 | -0.315 | 1.0×10^{-10} | 0.141 | 0.007 |
| 29 | -0.320 | 5.6×10^{-11} | 0.143 | 0.006 |
| 30 | -0.240 | 4.7×10^{-7} | 0.171 | 4.1×10^{-4} |
| 31 | -0.260 | 5.6×10^{-8} | 0.176 | 4.0×10^{-4} |
| 32 | -0.194 | 9.1×10^{-5} | 0.128 | 0.009 |
| 33 | -0.287 | 9.1×10^{-9} | -0.017 | 0.730 |
| 34 | -0.246 | 9.4×10^{-7} | 0.056 | 0.269 |
| 35 | -0.225 | 1.3×10^{-5} | -0.035 | 0.518 |
| 36 | -0.138 | 0.005 | 0.047 | 0.347 |
| 37 | 0.099 | 0.060 | 0.040 | 0.448 |
| 38 | 0.275 | 2.4×10^{-7} | 0.015 | 0.776 |
| 39 | 0.079 | 0.159 | -0.085 | 0.122 |
| 40 | 0.036 | 0.527 | -0.167 | 0.002 |

Table 9 Eukaryal versus Bacterial inter-domain analysis utilising the 5PR strategy

| | Nucleotides removed at the 5'-end of the 5PH 1 | Direct duplication model | | Inverse duplication model | |
|----|--|--------------------------|----------------------|---------------------------|-----------------------|
| | | Z(I) | P* | Z(I) | P* |
| 0 | | -0.157 | 0.002 | 0.014 | 0.767 |
| 1 | | -0.157 | 0.002 | 0.014 | 0.763 |
| 2 | | -0.157 | 0.002 | 0.014 | 0.765 |
| 3 | | -0.129 | 0.011 | -0.034 | 0.478 |
| 4 | | -0.121 | 0.016 | -0.033 | 0.495 |
| 5 | | -0.091 | 0.075 | -0.035 | 0.480 |
| 6 | | -0.061 | 0.223 | 0.019 | 0.688 |
| 7 | | 0.040 | 0.418 | 0.048 | 0.324 |
| 8 | | 0.045 | 0.365 | 0.050 | 0.302 |
| 9 | | 0.012 | 0.812 | 0.074 | 0.132 |
| 10 | | 0.153 | 0.002 | 0.149 | 0.001 |
| 11 | | 0.097 | 0.051 | 0.152 | 0.002 |
| 12 | | -0.238 | 1.1×10^{-6} | 0.149 | 0.002 |
| 13 | | -0.183 | 2.3×10^{-4} | 0.194 | 7.4×10^{-5} |
| 14 | | -0.183 | 5.0×10^{-4} | 0.293 | 3.7×10^{-9} |
| 15 | | -0.079 | 0.109 | 0.256 | 7.5×10^{-8} |
| 16 | | 0.064 | 0.192 | 0.292 | 1.5×10^{-9} |
| 17 | | 0.160 | 0.001 | 0.242 | 6.2×10^{-7} |
| 18 | | 0.128 | 0.008 | 0.227 | 2.9×10^{-6} |
| 19 | | 0.091 | 0.072 | 0.270 | 4.3×10^{-8} |
| 20 | | 0.190 | 9.4×10^{-5} | 0.217 | 6.8×10^{-6} |
| 21 | | 0.175 | 1.4×10^{-4} | 0.218 | 7.9×10^{-6} |
| 22 | | 0.163 | 4.7×10^{-4} | 0.224 | 4.9×10^{-6} |
| 23 | | 0.028 | 0.565 | 0.191 | 3.5×10^{-5} |
| 24 | | 0.029 | 0.541 | 0.189 | 4.1×10^{-5} |
| 25 | | -0.112 | 0.013 | 0.229 | 9.8×10^{-7} |
| 26 | | -0.111 | 0.014 | 0.229 | 9.9×10^{-7} |
| 27 | | -0.106 | 0.021 | 0.159 | 5.2×10^{-4} |
| 28 | | -0.289 | 7.1×10^{-9} | 0.259 | 9.8×10^{-8} |
| 29 | | -0.287 | 8.6×10^{-9} | 0.260 | 8.4×10^{-8} |
| 30 | | -0.242 | 2.6×10^{-7} | 0.339 | 5.2×10^{-13} |
| 31 | | -0.096 | 0.058 | 0.328 | 9.6×10^{-13} |
| 32 | | 0.066 | 0.181 | 0.338 | 2.1×10^{-12} |
| 33 | | -0.077 | 0.130 | 0.271 | 7.5×10^{-9} |
| 34 | | 0.177 | 5.6×10^{-4} | 0.331 | 1.6×10^{-11} |
| 35 | | 0.042 | 0.416 | 0.280 | 9.5×10^{-9} |
| 36 | | -0.027 | 0.611 | 0.198 | 3.3×10^{-5} |
| 37 | | 0.029 | 0.545 | 0.209 | 3.0×10^{-5} |
| 38 | | -0.208 | 1.3×10^{-5} | 0.204 | 7.9×10^{-5} |
| 39 | | -0.283 | 1.3×10^{-9} | 0.254 | 1.4×10^{-6} |
| 40 | | -0.045 | 0.333 | 0.210 | 2.1×10^{-4} |

The comparison of Bacterial and Archaeal domains confirms the trend observed in all other cases (Table 10) and this is in agreement to predictions of the inverse duplication model (Di Giulio 2010). On the contrary, the direct duplication model does not receive, as already seen, a strong support by this analysis (Table 10).

Finally, in Table 11 we summarise the statistical significance observed in the intra-molecular, inter-molecular and inter-domain comparisons, for the three strategies of nucleotide removal. In the inter-domain comparison we performed also another check. Besides performing, for instance, the Eukarya versus Archaea comparison, we also performed the

Table 10 Bacterial versus Archaeal inter-domain analysis utilising the 5PR strategy

| Nucleotides removed at the 5'-end of the 5PH | Direct duplication model | | Inverse duplication model | |
|--|--------------------------|----------------------|---------------------------|-----------------------|
| | Z(I) | P* | Z(I) | P* |
| 0 | -0.139 | 0.061 | 0.416 | 6.1×10^{-8} |
| 1 | -0.137 | 0.064 | 0.417 | 5.5×10^{-8} |
| 2 | -0.140 | 0.058 | 0.418 | 5.0×10^{-8} |
| 3 | -0.134 | 0.070 | 0.422 | 3.8×10^{-8} |
| 4 | -0.145 | 0.051 | 0.423 | 3.4×10^{-8} |
| 5 | -0.140 | 0.058 | 0.419 | 4.4×10^{-8} |
| 6 | -0.142 | 0.053 | 0.458 | 3.5×10^{-9} |
| 7 | -0.133 | 0.084 | 0.456 | 1.1×10^{-8} |
| 8 | -0.184 | 0.017 | 0.408 | 4.3×10^{-7} |
| 9 | -0.225 | 0.003 | 0.504 | 1.2×10^{-10} |
| 10 | -0.233 | 0.002 | 0.506 | 1.1×10^{-10} |
| 11 | -0.235 | 0.002 | 0.557 | 4.0×10^{-12} |
| 12 | -0.233 | 0.002 | 0.558 | 3.8×10^{-12} |
| 13 | -0.081 | 0.287 | 0.618 | 3.5×10^{-13} |
| 14 | 0.010 | 0.885 | 0.588 | 2.4×10^{-13} |
| 15 | -0.109 | 0.151 | 0.418 | 2.5×10^{-7} |
| 16 | -0.180 | 0.006 | 0.407 | 1.1×10^{-6} |
| 17 | -0.190 | 0.004 | 0.337 | 2.0×10^{-4} |
| 18 | -0.214 | 0.002 | 0.296 | 5.1×10^{-4} |
| 19 | -0.165 | 0.017 | 0.455 | 2.7×10^{-7} |
| 20 | -0.195 | 0.004 | 0.490 | 1.8×10^{-9} |
| 21 | -0.181 | 0.006 | 0.489 | 1.8×10^{-9} |
| 22 | -0.296 | 9.8×10^{-6} | 0.381 | 5.6×10^{-6} |
| 23 | -0.324 | 1.4×10^{-6} | 0.416 | 1.6×10^{-6} |
| 24 | -0.324 | 1.4×10^{-6} | 0.417 | 1.5×10^{-6} |
| 25 | -0.261 | 2.1×10^{-4} | 0.482 | 7.6×10^{-8} |
| 26 | -0.446 | 1.2×10^{-9} | 0.388 | 5.9×10^{-6} |
| 27 | -0.290 | 2.2×10^{-4} | 0.406 | 3.8×10^{-7} |
| 28 | -0.290 | 2.2×10^{-4} | 0.406 | 3.8×10^{-7} |
| 29 | -0.308 | 6.0×10^{-5} | 0.256 | 8.7×10^{-4} |
| 30 | -0.331 | 2.6×10^{-5} | 0.230 | 0.002 |
| 31 | -0.305 | 8.5×10^{-5} | 0.223 | 0.003 |
| 32 | -0.305 | 8.5×10^{-5} | 0.223 | 0.003 |
| 33 | -0.352 | 5.5×10^{-6} | 0.240 | 9.7×10^{-4} |
| 34 | -0.358 | 2.8×10^{-6} | 0.240 | 9.6×10^{-4} |
| 35 | -0.184 | 0.018 | 0.280 | 6.3×10^{-5} |
| 36 | -0.114 | 0.175 | 0.161 | 0.023 |
| 37 | -0.345 | 2.6×10^{-5} | 0.046 | 0.503 |
| 38 | -0.388 | 8.1×10^{-7} | -4.8×10^{-4} | 0.994 |
| 39 | -0.387 | 8.3×10^{-7} | -0.005 | 0.946 |
| 40 | -0.446 | 6.2×10^{-9} | 0.024 | 0.715 |

Archaea versus Eukarya comparison, which differed for the domain to which belongs the 5S rRNA molecule from which nucleotides were removed. Results are in agreement with the indirect duplication model and against the direct one (Table 11).

Summary and General Discussion

We have tested the hypotheses that the origin of the 5S rRNA molecule could have occurred through a direct or indirect duplication. The results, here reported, support a hypothesis

Table 11 Summary of results obtained in the intra-molecular, inter-molecular (A) and inter-domain (B) comparisons, using the 5PR strategy, RR strategy and 3PR strategy (see ‘Results’)

| | 5PR | | | | RR | | | | 3PR | | | |
|-----------------------|-----------------|----|-----------------|----|-----------------|----|-----------------|----|-----------------|---|-----------------|----|
| | Intra-molecular | | Inter-molecular | | Intra-molecular | | Inter-molecular | | Intra-molecular | | Inter-molecular | |
| | D | I | D | I | D | I | D | I | D | I | D | I |
| (A) | | | | | | | | | | | | |
| Eucaryal | – | ++ | – | ++ | – | ++ | – | ++ | + | – | + | ++ |
| Archaeal | -- | ++ | – | ++ | – | ++ | – | ++ | ± | – | ± | + |
| Bacterial | – | ++ | – | ++ | – | ++ | – | ++ | – | – | – | ++ |
| | 5PR | | | | RR | | | | 3PR | | | |
| | D | | I | | D | | I | | D | | I | |
| (B) | | | | | | | | | | | | |
| Eucaryal vs Archaeal | | – | | ++ | + | | ++ | – | | + | | |
| Bacterial vs Eucaryal | | – | | ++ | – | | ++ | ++ | | + | | |
| Archeal vs Bacterial | | – | | ++ | ++ | | ++ | – | | – | | |
| Archaeal vs Eucaryal | | – | | ++ | – | | + | + | | + | | |
| Eucaryal vs Bacterial | | – | | ++ | – | | ++ | – | | – | | ++ |
| Bacterial vs Archaeal | | – | | ++ | – | | ++ | – | | – | | + |

Results are reported for the direct (D) and inverse (I) duplication models. The ++, +, – indicate that the analysis is highly, marginally, not statistical significant, respectively

of complementarity between the two halves of the 5S RNA molecule (Tables 2, 3, 4, 5, 6, 7, 8, 9, 10, 11), favouring the indirect duplication model (Fig. 1; Di Giulio 2010).

In particular, the indirect duplication model is supported by the intra-molecular analysis in all three domains of life (Fig. 3; Tables. 2, 11; see ‘Results’), whereas the direct duplication model does not receive an adequate corroboration (Tables 2, 4, 11). Moreover, the inter-molecular analysis confirms the intra-molecular one, corroborating again the inverse duplication model and not the direct one (Tables 5, 6, 7, 11; Fig. 4). Also, the severe test represented by the inter-domain comparison favours the inverse duplication model and not the direct one (Tables 8, 9, 10, 11; see ‘Results’). In particular, the comparison between 5S rRNA molecules from Eucaryal and Bacterial domains, showing the highest divergence in the secondary structure (Fig. 2), favours again the indirect duplication model and not the direct one (Tables 9, 11).

Evidently, also the total statistical significance of the indirect model is higher than that of the direct model (Table 11). However, there is an exception in the intra-molecular analysis (using the 3PR strategy), in which the indirect model did not show a statistical significance (Table 11A). We interpret this curious observation and cases in which the direct model shows some degree significance by assuming that this might be either (i) an indication that the removal of nucleotides favours the right alignment between the two halves of the 5S rRNA molecule, or (ii) reflecting the hypothesis that a hairpin-like molecule was the precursor of

these duplications. Thus, in this case it is impossible to distinguish between the indirect and the direct models, as already we have previously and extensively discussed (Branciamore and Di Giulio 2011).

More generally, these observations along with the ones made for other RNA molecules showing symmetry or semi-symmetry (Di Giulio 1992, 2010; Bokov and Steinberg 2009), corroborate the hypothesis, here supported, that in the case a molecule of RNA or protein shows a symmetry, this must necessarily have been generated by means of gene duplication and not by a gradualistic accretion. Thus, we reject those models that try to explain the evolution of some symmetrical molecules by means of gradualistic accretion (Sun and Caetano-Anollés 2008, 2010; Rodin et al. 2011) and instead noting the extraordinary capability the nucleic acids have of duplicating themselves, we believe the most parsimonious way of generating these symmetries should be used.

Acknowledgments Part of this work has been written at the Allan Wilson Centre for Molecular Ecology and Evolution during a stay of MDG at the Massey University (Palmerston North, New Zealand). I would like to thank David Penny, Peter Lockhart and Ibrar Ahmed for their discussions and cordial hospitality. Finally, we are pleased to thank two anonymous and patient reviewers.

References

Bokov K, Steinberg SV (2009) A hierarchical model for evolution of 23S ribosomal RNA. *Nature* 457:977–980

- Branciamore S, Di Giulio M (2011) The presence in tRNA molecule sequences of the double hairpin, an evolutionary stage through which the origin of this molecule is thought to have passed. *J Mol Evol* 72:352–363
- Chamary JV, Hurst LD (2005) Evidence for selection on synonymous mutations affecting stability of mRNA secondary structure in mammals. *Genome Biol* 6:R75
- Di Giulio M (1992) On the origin of the transfer RNA molecule. *J Theor Biol* 159:199–214
- Di Giulio M (2004) The origin of the tRNA molecule: implications for the origin of protein synthesis. *J Theor Biol* 226:89–93
- Di Giulio M (2006) The non-monophyletic origin of the tRNA molecule and the origin of genes only after the evolutionary stage of the Last Universal Common Ancestor (LUCA). *J Theor Biol* 240:343–352
- Di Giulio M (2009) A comparison among the models proposed to explain the origin of the tRNA molecule: a synthesis. *J Mol Evol* 69:1–9
- Di Giulio M (2010) A model of the origin of the 5S ribosomal RNA molecule. *J Mol Evol* 71:1–2
- Edgar RC (2004) MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics* 5:113
- Fox GE, Woese CR (1975) 5S RNA secondary structure. *Nature* 256:505–507
- Hannock J, Wagner R (1982) A structural model of 5S RNA from *E. coli* based on intramolecular crosslinking evidence. *Nucleic Acids Res* 10:1257–1269
- Joachimiak A, Nalaskowska N, Barciszewska M, Barciszewski J, Mashkova T (1990) Higher plant 5S rRNAs share common secondary and tertiary structure. A new three domains model. *Int J Macromol* 12:321–327
- Luehrsen KR, Fox GE (1981) Secondary structure of eukaryotic cytoplasmic 5S ribosomal RNA. *Proc Natl Acad Sci USA* 78:2150–2154
- Ohno S (1970) *Evolution by gene duplication*. Springer, Berlin
- Rodin AS, Szathmáry E, Rodin SN (2011) On origin of genetic code and tRNA before translation. *Biol Direct* 6:14
- Sun F-J, Caetano-Anolles G (2008) The origin and evolution of tRNA inferred from phylogenetic analysis of structure. *J Mol Evol* 66:21–35
- Sun F-J, Caetano-Anolles G (2009) The evolutionary history of the structure of 5S ribosomal RNA. *J Mol Evol* 69:430–443
- Sun FJ, Caetano-Anollés G (2010) The origin of modern 5S rRNA: a case of relating models of structural history to phylogenetic data. *J Mol Evol* 71:3–5
- Szymanski M, Barciszewska MZ, Barciszewski J, Erdmann VA (2000) 5S ribosomal RNA database Y2K. *Nucl Acids Res.* 28:166–167. <http://biobases.ibch.poznan.pl/5SData/>