

Botanisches Institut, Karlsruher Institut für Technologie

rDNA Marker

Anwendung in der Evolutionsforschung

Thomas Horn
19.05.2014

Inhalt

I.	Ribosomen.....	1
II.	Organisation der Ribosomalen DNA in Eukaryoten	3
III.	Verwendung von rDNA in Phylogenetischen Analysen.....	4
IV.	„Concerted Evolution“	7
V.	Zusammenfassung.....	9
VI.	Primer	9
A.	ITS: ITS5 / ITS4	10
B.	ITS: ITS-A / ITS-B	10
C.	ITS: 18F / 28R.....	11
D.	ITS2: BEL1 – BEL4.....	11
E.	NTS: S1 / AS1	11
F.	NTS: 5SP1 / 5SP2	12
VII.	Abkürzungen.....	12
VIII.	Referenzen	13

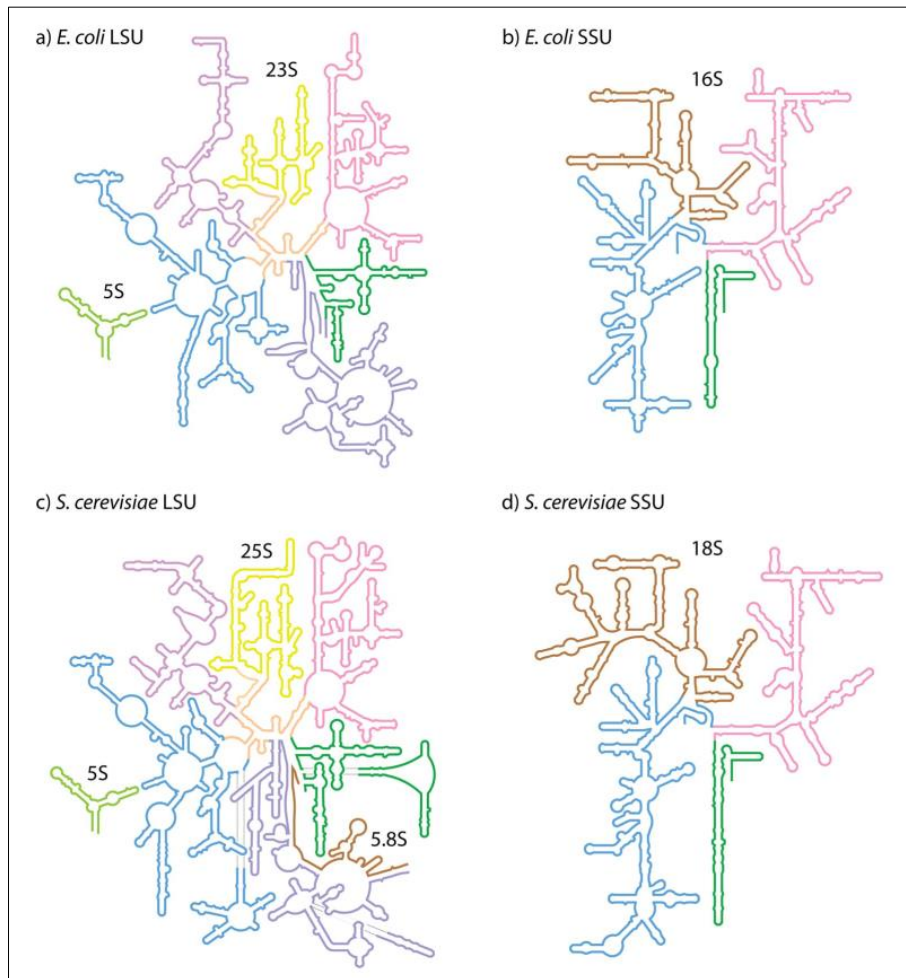
Notiz:

Verbesserungsvorschläge und Fragen jeglicher Art sind willkommen!

I. Ribosomen

Ribosomen sind Komplexe aus Proteinen und RNA Molekülen (ribosomale RNA), die eine zentrale Rolle bei der Proteinbiosynthese in allen zellulären Lebensformen spielen. Sie bestehen aus zwei Untereinheiten, der kleinen (SSU), welche die messenger RNA Matrix liest und die große (LSU), welche die entsprechenden Aminosäuren verkettet. Hierbei liefert die transfer RNA – ein relativ kleines RNA Molekül (4 S) – die entsprechenden Aminosäuren zu den Ribosomen. Verschiedene RNA Moleküle (**Abbildung 1**) bilden zusammen mit Proteinen die Ribosomen.

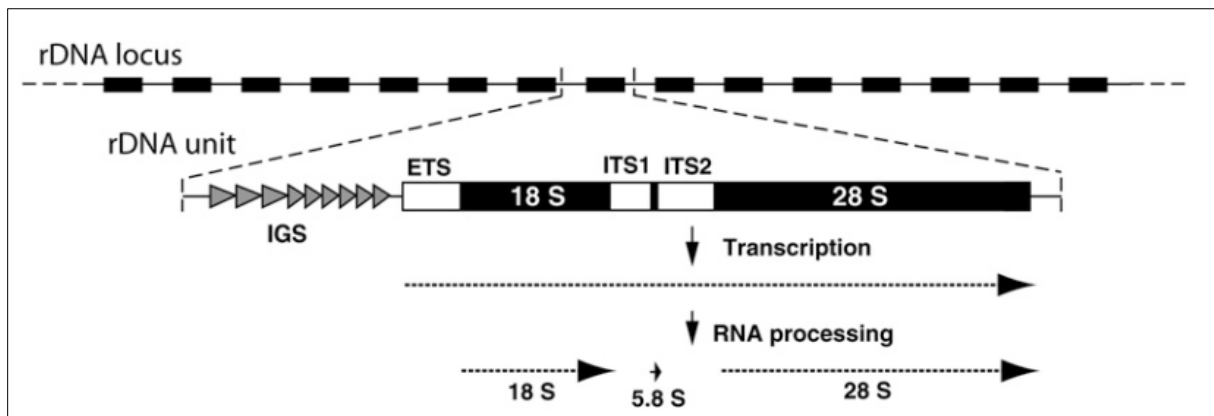
Abbildung 1: Schematische 2-D Strukturen der ribosomale RNA von a) *E. coli* LSU, b) *E. coli* SSU, c) *S. cerevisiae* LSU, and d) *S. cerevisiae* SSU. [Petrov 2014 Fehler! Verweisquelle konnte nicht gefunden werden.]



Die Information zur Synthese der ribosomalen RNA (rRNA) liegt, wie die der Proteine, kodiert in der DNA vor. Eine Vorstufe wird durch RNA Polymerase synthetisiert und in weiteren Schritten zur reifen rRNA prozessiert.

II. Organisation der Ribosomalen DNA in Eukaryoten

Abbildung 2: Schematische Darstellung der Organisation der ribosomalen DNA (rDNA) [Eickbush 2007]. Im oberen Teil ist ein rDNA Locus dargestellt in dem sich die rDNA Einheiten (schwarze Rechtecke) – getrennt durch „intergenic spacer“ (IGS) bzw. „non-transcribed spacer“ (NTS) – mehrfach wiederholen. Darunter ist der Aufbau einer rDNA Einheit dargestellt und im Folgenden das Transkript sowie die prozessierte ribosomale RNA (rRNA).

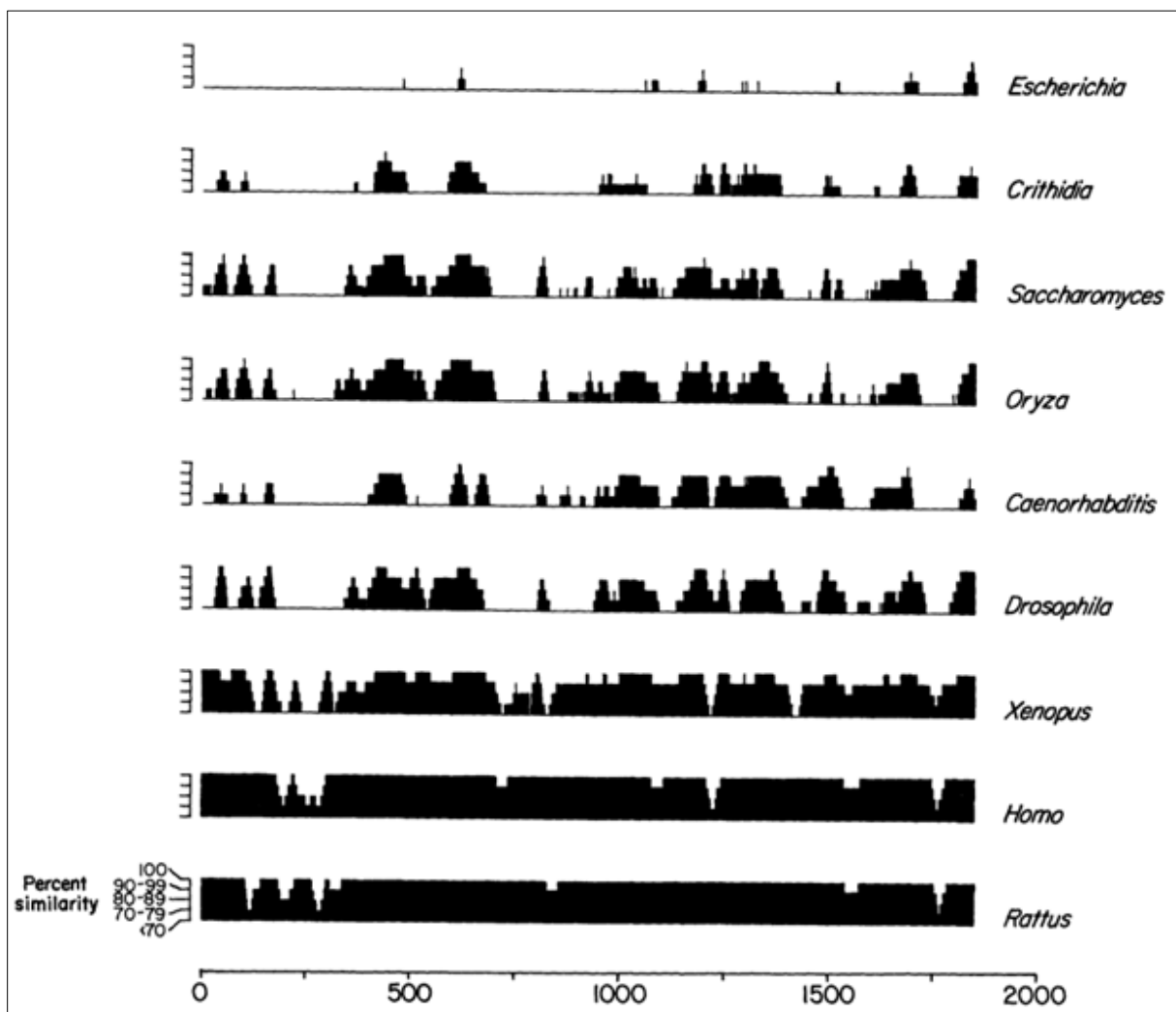


Charakteristisch für die meisten eukaryotischen Lebensformen ist die Organisation der rDNA der kleinen Untereinheit (18 S) und zwei der großen Untereinheit (5.8 S und 28 S) in einer Transkriptionseinheit (**Abbildung 2**). Diese beginnt mit dem externen Spacer (ETS) und die darauf folgenden rDNA Bereiche sind durch zwei interne Spacer (ITS1 und ITS2) voneinander getrennt. Wie in **Abbildung 2** zu sehen, gibt es an einem Locus eine Vielzahl solcher Einheiten die tandemartig hintereinander aufgereiht und durch sogenannte „intergenic spacer“ (IGS) voneinander getrennt sind. In Harmonie mit der Benennung der transkribierten Spacer der ribosomalen DNA spricht man auch von „non-transcribed spacer“ (NTS) anstatt IGS.

Eine an anderer Stelle lokalisierte Transkriptionseinheit beinhaltet die 5 S rDNA (**Abbildung 10** und **Abbildung 11**) und kommt wie die oben beschriebene in tandemartigen Wiederholungen welche durch NTS voneinander getrennt sind im Genom aller Lebensformen vor.

III. Verwendung von rDNA in Phylogenetischen Analysen

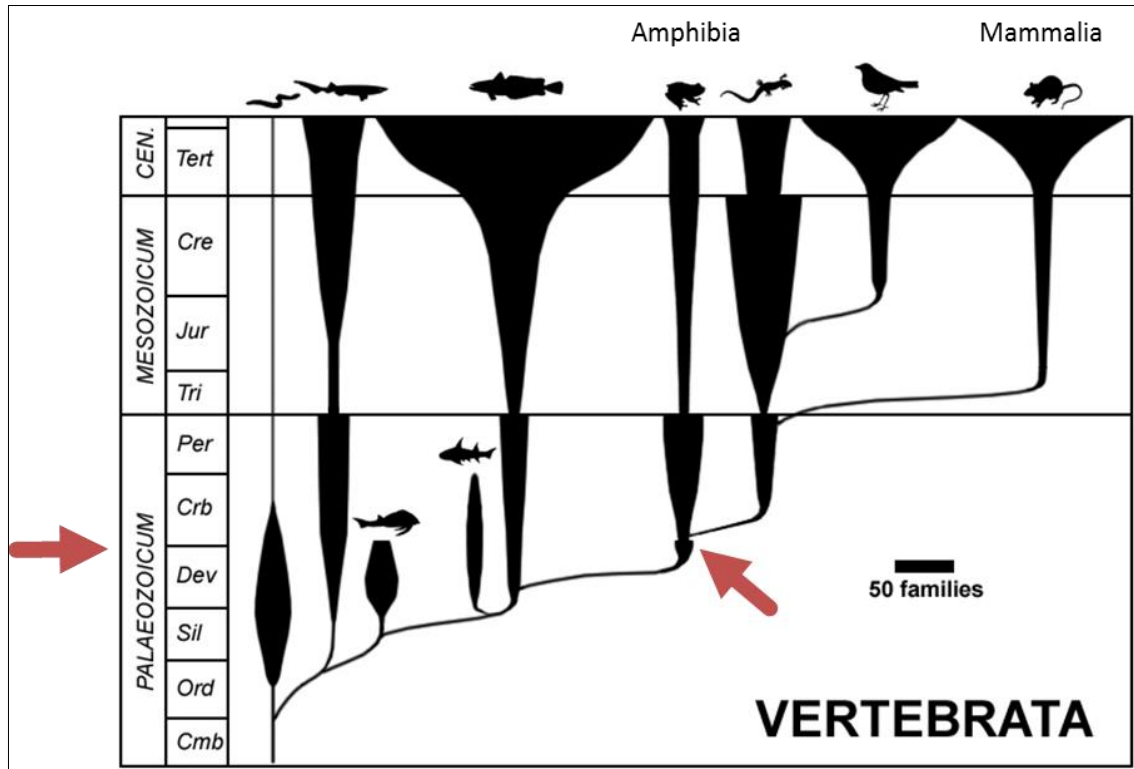
Abbildung 3: Sequenz Vergleich des rRNA Gens der kleinen Ribosomen Untereinheit (16 S – 18 S) [Hillis 1991]. Die rDNA Sequenz der Hausmaus (*Mus musculus*, Mammalia, Vertebrata) wurde jeweils mit der homologen Sequenz von *Escherichia coli* (Eubakterium), *Crithidia fasciculata* (Trypanosoma, Protozoa), *Saccharomyces cerevisiae* (Fungi, Ascomycota), *Oryza sativa* (Poaceae, Planta), *Caenorhabditis elegans* (Nematoda, Invertebrata), *Drosophila melanogaster* (Diptera, Invertebrata), *Xenopus laevis* (Amphibia, Vertebrata), *Homo sapiens* (Mammalia, Vertebrata) und *Rattus norvegicus* (Mammalia, Vertebrata) verglichen. Die Analyse wurde mit dem sogenannten „Sliding-Window“ Verfahren erstellt. Dabei wird die Ähnlichkeit zweier Sequenzen Stückweise („Window“ bezieht sich hier auf einen Teil-Bereich der Gesamtsequenz) bestimmt. In dieser Analyse wurde ein Bereich von 30 Basen Paaren (bp) verglichen (Position 1 - 30), das Fenster um 4 bp verschoben und erneut verglichen (Position 5 - 35). Dieses Vorgehen wird wiederholt bis das Ende der Sequenz erreicht ist. In der Abbildung werde die Ähnlichkeitswerte für den Bereich zwischen 70 und 100 % pro „Window“ angezeigt.



Die am besten studierte rDNA ist die der kleinen Untereinheit (16 S - 18 S). Sie hat sich nur wenig während der letzten 300 – 400 Millionen Jahre verändert. In **Abbildung 3** ist das Ergebnis eines Sequenzvergleichs der rDNA der kleinen Untereinheit dargestellt (Details zur Analyse stehen in der Beschreibung der Abbildung). Als Referenz Sequenz diente die rDNA der Hausmaus (*Mus musculus*). Im Vergleich zur rDNA der Ratte (*Rattus norvegicus*) sowie des Menschen (*Homo sapiens*) findet man nur wenige variable Bereiche (weiß) und selbst bei den Amphibien (*Xenopus laevis*) ist der größte Teil noch stark konserviert (schwarz).

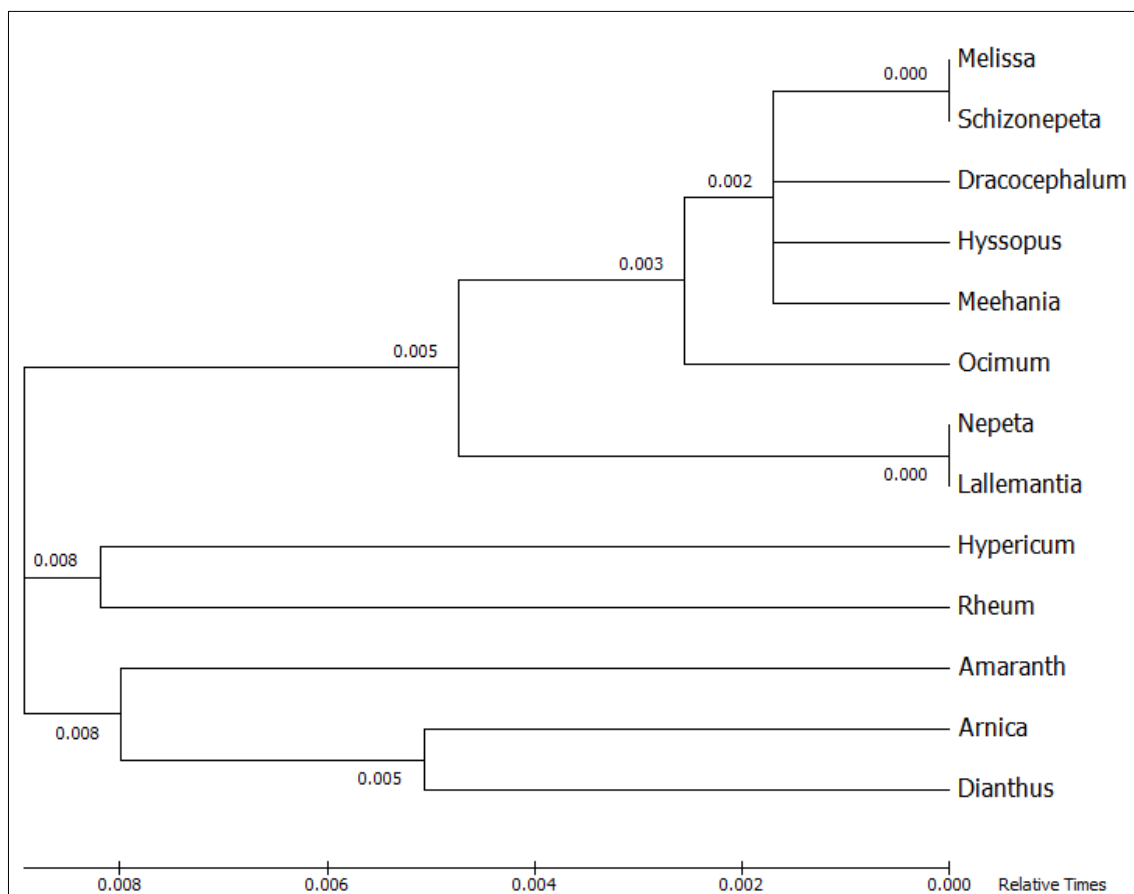
Aus diesem Ergebnis kann man ableiten, dass sich das rRNA Gen der kleinen Untereinheit nur für Untersuchungen von Gruppen eignet, die sich bereits mehr als 350 Millionen Jahre unabhängig voneinander entwickelt haben.

Abbildung 4: Evolution der Wirbeltiere vom Cambrium bis zur heutigen Periode (Quaternary) [Benton 1998]. Die Entstehung der Wirbeltiere Klassen ist in Form eines Spindeldiagramms dargestellt. Die Breite einer Spindel repräsentiert die Vielfalt innerhalb einer Klasse hinsichtlich der Anzahl an bekannten Familien. Die roten Pfeile weisen auf die Abspaltung einer Gruppe hin (vor ca. 360 Millionen Jahren), aus der sich im Verlauf der Evolution unter anderem die Säugetiere (*Mammalia*) entwickelt haben.



Die rDNA der großen Untereinheit (23 S - 28 S) ist länger und verändert sich etwas schneller als die der kleinen. Aufgrund ihrer wesentlich kürzeren Sequenzlänge werden 5.8 S und 5 S rDNA eher selten für Untersuchungen verwendet.

Abbildung 5: Evolutionäre Beziehung von ausgewählten Pflanzen Gattungen basierend auf der Neighbor-Joining Methode. Für die Analyse wurden 5.8 S rDNA Sequenzen verwendet. Die Analyse basiert auf 13 Nucleotid Sequenzen mit 165 Positionen. Die Gruppe im oberen Teil besteht aus Gattungen der Ordnung *Lamiales* (*Lamiaceae*: *Melissa*, *Schizonepeta*, *Dracocephalum*, *Hyssopus*, *Meehania*, *Ocimum*, *Nepeta* und *Lallemantia*). Die verbleibenden Gattungen gehören zur Ordnung *Caryophyllales* (*Rheum*, *Amaranthus* und *Dianthus*), *Malpighiales* (*Hypericum*) und *Asterales* (*Arnica*). Der Baum wurde mit MEGA6 erstellt [Tamura 2013].



Neben den erwähnten kodierenden Bereichen der rDNA werden auch die wesentlich variablen (also weniger konservierten) Bereiche der Spacer (NTS, ETS und ITS) für Phylogenetische Analysen verwendet. Diese eignen sich am besten für die Untersuchung von relativ jungen Aufspaltungen. Will man z.B. für den in **Abbildung 5** dargestellten Vergleich ITS Sequenzen verwenden stellt man fest, dass es aufgrund zu hoher Divergenz zwischen den Gattungen zu keinem verwendbaren Alignment kommt. Wenn man sich jedoch auf die

Familie der Lippenblütler (Lamiaceae) beschränkt erhält man ein brauchbares Alignment und kann eine Analyse durchführen. Im Vergleich zwischen NTS, ETS und ITS erwiesen sich die NTS als die sich am schnellsten verändernden Spacer.

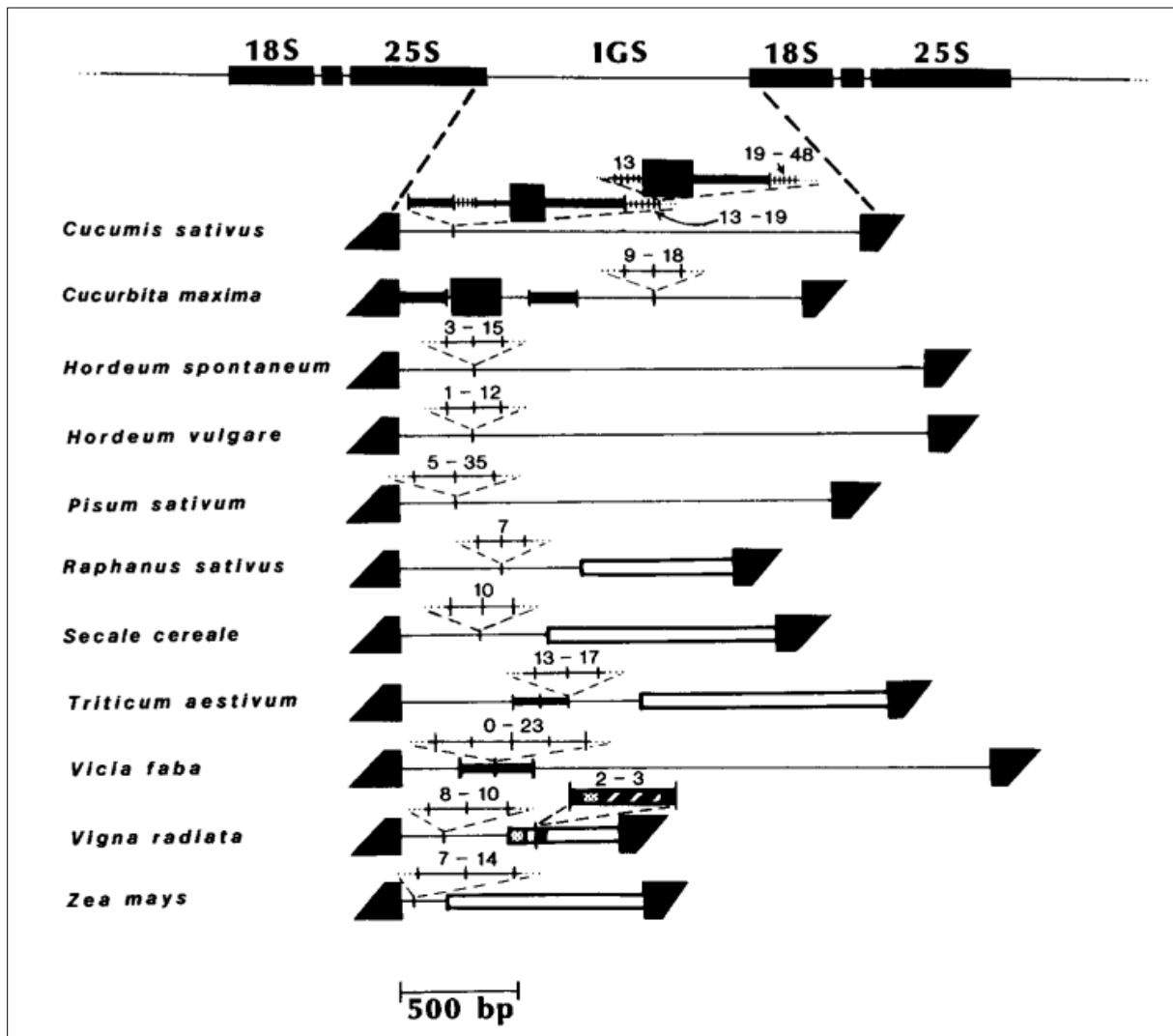
IV. „Concerted Evolution“

In der Regel verändern sich Bereiche des Genoms unterschiedlich schnell. Je wichtiger der Locus, umso weniger zugänglich ist dieser für Veränderungen. Individuen mit einer Mutation in diesem Bereich tragen als Konsequenz nur wenig oder gar nicht zum Bestand der Population bei. Daher etabliert sich diese Variante erst gar nicht und verschwindet im Gefüge der Zeit. Mutationen in Bereichen, die kaum einen Effekt auf die Lebensfähigkeit und das Fortpflanzungspotential haben, können sich dahingegen länger halten, und bei entsprechendem Erfolg des Individuums und seiner Nachkommen, innerhalb einer Population etablieren.

Wenn eine Kopie eines notwendigen Locus durch Duplikation entsteht (Paralog) kann theoretisch eine der Varianten mutieren ohne dass ein Nachteil daraus resultiert. Auf diese Weise entstehen sogenannte Pseudogene (Gene die ihre Funktion verloren haben) und unter Umständen neue Varianten des ursprünglichen Gens, welche abgewandelte Funktionen erfüllen können.

Die in Abbildung 2 dargestellte Transkriptionseinheit taucht in vielen Organismen Gruppen in hoher Kopien Zahl auf. Man könnte nun erwarten, dass diese Kopien sich unabhängig voneinander entwickeln und wie oben beschrieben Pseudogene oder neue Varianten entstehen. Studien der rDNA zeigen kaum Unterschiede in den rDNA Bereichen innerhalb einer Art, was darauf hinweist, dass sich diese nicht unabhängig voneinander entwickeln. Es kommt vielmehr zu einer Homogenisierung der verschiedenen Kopien („concerted evolution“) [[Hillis 1991](#)]. Mechanismen die in diesem Zusammenhang diskutiert werden sind „gene conversion“ und „unequal crossing over“ [[Eickbush 2007](#)]. Die häufigsten Variationen die innerhalb einer Art beobachtet werden können, sind Unterschiede innerhalb der NTS und in der Anzahl von wiederholten Transkriptionseinheiten.

Abbildung 6: Schematischer Vergleich von 25S-18S-NTS Regionen (=IGS) verschiedener Pflanzen Taxa [Rogers 1987]. Es wird veranschaulicht, dass sich die NTS zwischen den Taxa deutlich unterscheiden. Zum einen in der Länge und zum anderen durch die beschriebenen Strukturen. Beispielsweise sieht man in *Raphanus sativus*, dass 7 repetitive Elemente und ein ETS (weißes Rechteck) vorkommen. In *Pisum sativum* wurden mehrere Individuen untersucht und führte zu der Angabe 5 – 35 für die Anzahl der repetitiven Elemente. In *Cucurbita maxima* findet man eine Kopie aus der 25S rDNA in der Spacer Region (großes schwarzes Rechteck) und zwei verschiedene Arten von repetitiven Elementen (schmale schwarze Rechtecke und die 9 – 18 Elemente).



V. Zusammenfassung

Die ribosomale DNA bietet verschiedene Marker an, mit denen evolutionäre Fragen von unterschiedlicher zeitlicher Tiefe adressieren werden können. Mit Informationen der stark konservierten kodierenden Bereiche können die Wurzeln der verschiedenen Lebensformen verglichen werden. Abweichende Varianten sowie kodierende Bereiche der Organellen rRNA können neuere Verzweigungen im Baum des Lebens beleuchten. Mit den Spacer Abschnitten ist es möglich innerhalb von Familien, Gattungen und Arten die Verwandtschaften aufzuklären.

Auch wenn die Kopien der Transkriptionseinheiten einem homogenisierenden Mechanismus unterliegen, sollte man die mögliche Existenz von Paralogen nicht ignorieren [Buckler 1997].

VI. Primer

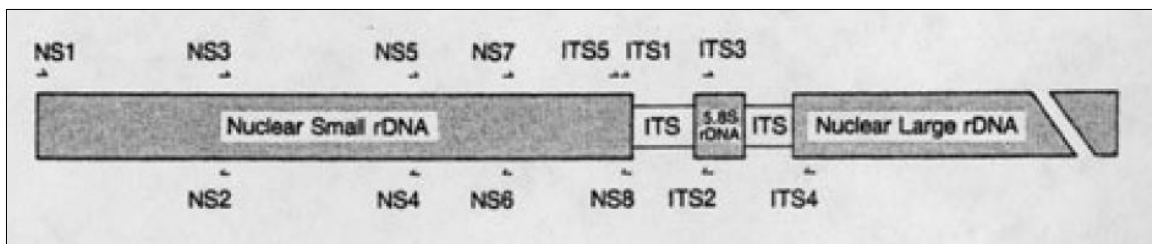
Tabelle 1: Primer zur Amplifikation von Spacer Regionen der ribosomalen DNA. Der Begriff „Universal“ unter Referenz deutet auf die erprobte Anwendung in unterschiedlichen Taxa (Arten, Gattungen, Familien) hin. Es kann aber durchaus sein, dass die Primer nicht in allen Pflanzen Taxa (also wirklich universell) zu einer erfolgreichen Amplifikation führen. Wenn mehrere Banden auftauchen sollten weist dies entweder auf eine Kontamination oder auf kürzere/längere Versionen des entsprechenden Bereichs hin. Im Fall der NTS Primer ist damit am häufigsten zu rechnen, was auch der Grund für deren Einsatz als DNA Fingerprinting Primer ist.

Name		5'-3' Sequenz	Lokus	Referenz
ITS5	f	GGAAGTAAAAGTCGTAACAAGG	ITS1, 5.8 S, ITS2	"Universal" [White 1990]
ITS4	r	TCCTCCGCTTATTGATATGC		
18F	f	GTGAAGTGGGAAGGATCTTG	ITS1, 5.8 S, ITS2	"Universal" [Lu 2005, 2007]
28R	r	CCGCCTGACCTGRGGTC		
ITS-A	f	GGAAGGAGAAGTCGTAACAAGG	ITS1, 5.8 S, ITS2	"Universal" [Blattner 1999]
ITS-B	r	CTTTTCCTCCGCTTATTGATATG		
ITS-C		GCAATTCACACCAAGTATCGC	ITS1	
ITS-D		CTCTCGGCAACGGATATCTCG	ITS2	
BEL-1	f	GGDGC GGAKAHTGGCCYCCCGTGC	ITS2	"Universal" [Chiou 2007]
BEL-2	f	GATGCGGAGATTGGCCCCCGTGC		
BEL-3	r	GACGCTTCTCCAGACTACAAT		
BEL-4	r	GTTTCTTTTCCTCCGCT		
18S-IGS	r	GAGACAAGCATATGACTACTGGCAGGATCAACCAG	NTS + ETS	"Universal" (Euk) [Baldwin 1998]
26S-IGS	f	GGATTGTTCAACCACCAATAGGGAACGTGAGCTG		
18S-ETS	r	ACTTACACATGCATGGCTTAATCT	ETS	"Universal" (Ang)
A-ETS-F	f	TTCGTGGTYGGCGGAGCT		
NTSF	f	GATCCCATCAGAACTCCGAAG	5 S NTS	<i>Panax</i>

NTSR	r	CGGTGATTAGTGCTGGTAT		[Diao 2009]
NTS-5SP1	f	GTGCTTGGGCGAGAGTAGTA	5 S NTS	"Universal" [Mizukami 1995]
NTS-5SP2	r	TTAGTGCTGGTATGATCGCA		
NTS-S1	f	GGATCCGTGCTTGGGCGAGAGTAGTA	5 S NTS	"Universal" [Carles 2001]
NTS-AS1	r	GGATCCTTAGTGCTGGTATGATCGCA		

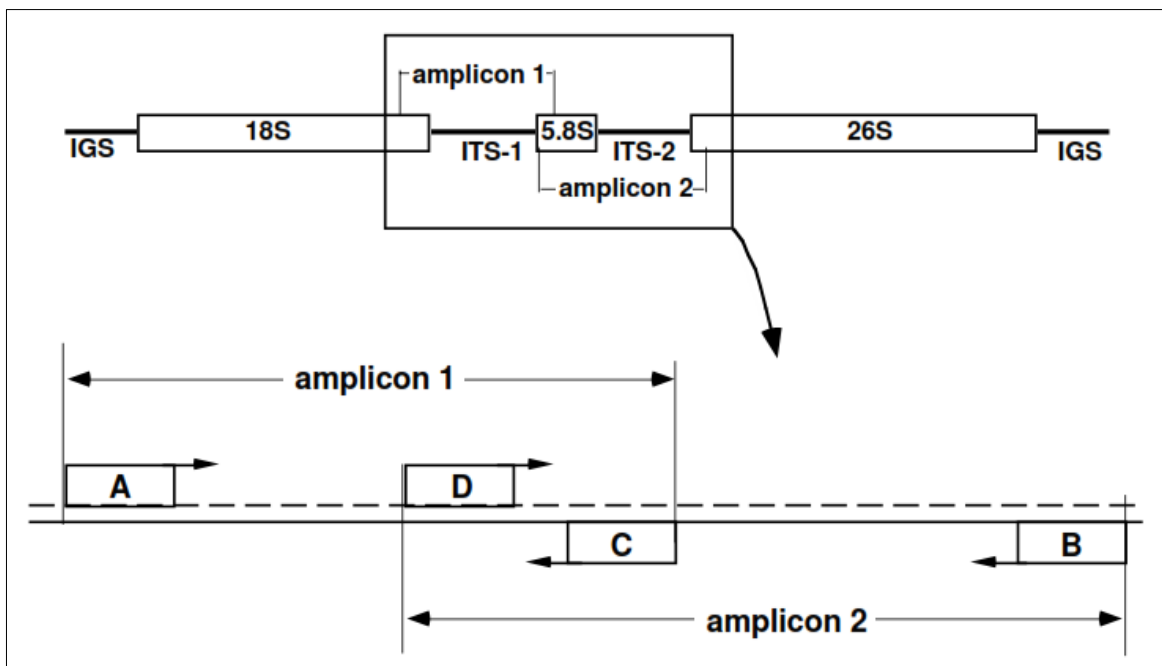
A. ITS: ITS5 / ITS4

Abbildung 7: Schema der eukaryotischen Transkriptionseinheit (siehe auch **Abbildung 2**) mit angedeuteten Primer Positionen [White 1990]. Die Primer wurden ursprünglich für die Amplifikation entsprechender Bereiche aus Pilzen entwickelt.



B. ITS: ITS-A / ITS-B

Abbildung 8: Schema der eukaryotischen Transkriptionseinheit (siehe auch **Abbildung 2**) mit angedeuteten Primer Positionen [Blattner 1999]. Diese Primer basieren auf denen von White [1990] und wurden von Blattner für die Amplifikation aus schwierigem pflanzlichem Ausgangsmaterial modifiziert.



C. ITS: 18F / 28R

Hinweise auf die Entwicklung der Primer 18F und 28R [Lu 2005, 2007] beschränken sich auf eine sehr generelle und somit nicht sehr hilfreiche Aussage der Autoren: „Die Primer basieren auf den konservierten Regionen der 18 S und 26 S rRNA Gene und der Wobble Theorie“.

D. ITS2: BEL1 – BEL4

Abbildung 9: Schema der eukaryotischen Transkriptionseinheit (siehe auch **Abbildung 2**) mit angedeuteten Primer Positionen [Chiou 2007]. Diese Primer wurden auf der Basis von Sequenzinformationen von 120 Arten aus 52 Familien entwickelt. Die Primer Kombination BEL-1 / BEL-3 führte zur erfolgreichen Amplifikation des entsprechenden Fragments bei allen 120 Arten. Die Kombination BEL-2 / BEL-3 erwies sich effektiver bei der Amplifikation unter Verwendung von Arten aus den Familien *Papaveraceae*, *Zingiberaceae*, *Liliaceae* und *Araceae*. Warum die Autoren allerdings die Primer BEL-1 und BEL-2 in die Spacer Region gelegt haben und nicht in den stärker konservierten Bereich der 5.8 S rDNA wird nicht erwähnt.

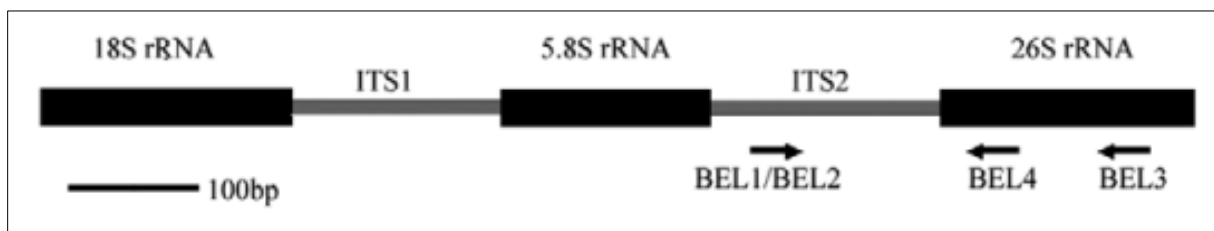


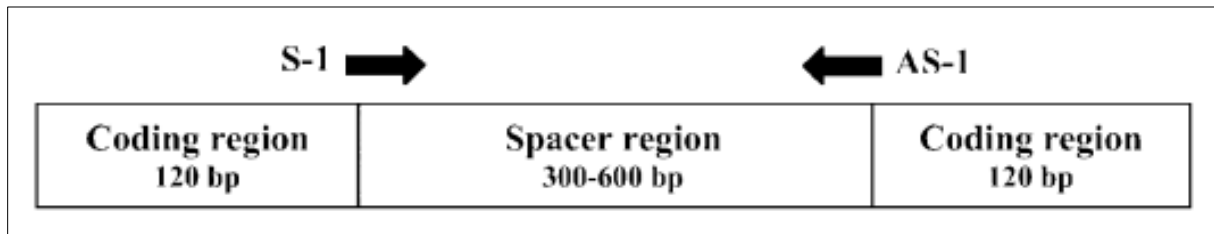
Tabelle 2: PCR Programm zur Amplifikation mit Primer Paar BEL-1 / BEL-3 und BEL-2 / BEL-3. Der Mastermix unterscheidet sich zu unserem Standard Mix in der finalen Magnesium Konzentration (1.5 mM statt 2 mM), der finalen dNTP Konzentration (150 µM statt 200 µM) und dem zusätzlichen Einsatz von DMSO (3 µl auf 50 µl).

Initial Denaturation	94° C	5 min
40 Cycles		
Denaturation	94° C	30 sec
Annealing	56° C	30 sec
Extension	68° C	45 sec
Cycle End		
Final Extension	68° C	10 min

E. NTS: S1 / AS1

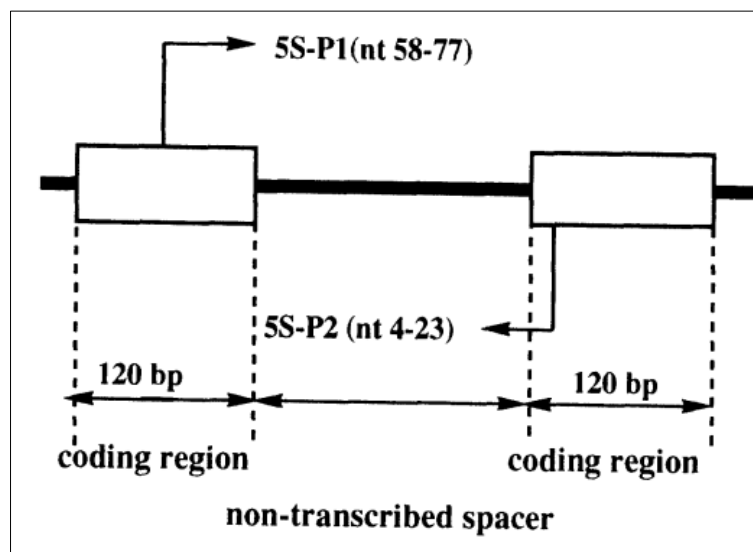
Abbildung 10: Schema der NTS Region der 5S-rDNA mit angedeuteten Primer Positionen [Carles 2001]. Wie genau es zur Entstehung dieser Primer kam ist unklar. Carles führt in dem

Zusammenhang eine Referenz an [\[Cai 1999\]](#) in welcher man auf eine weitere Referenz verwiesen wird [\[Mizukami 1995\]](#). In dieser werden allerdings die Primer 5S-P1 und 5S-P2 beschrieben und verwendet. Möglicherweise handelt es sich bei den Primer S1 und AS1 um eine Entwicklung von Cai die auf derselben Grundlage basiert wie Mizukamis [\[Wolters 1988\]](#). Da nur Mizukami eindeutig auf die Grundlage der Primer Entwicklung hinweist, scheint dies die plausibelste Erklärung zu sein.



F. NTS: 5SP1 / 5SP2

Abbildung 11: Schema der NTS Region der 5S-rDNA mit angedeuteten Positionen der Primer 5S-P1 und 5S-P2 [\[Mizukami 1995\]](#). Die Entwicklung dieser Primer basiert auf einer Kompilation von 5S-rDNA Sequenzen höherer Pflanzen [\[Wolters 1988\]](#).



VII. Abkürzungen

LSU - Large Subunit, große Untereinheit der Ribosomen

SSU	-	Small Subunit, kleine Untereinheit der Ribosomen
S	-	Svedberg, Maßeinheit des Sedimentationskoeffizienten
rRNA	-	ribosomale RNA
rDNA	-	ribosomale DNA (kodierte Version der rRNA)
IGS	-	Intergenic Spacer
ETS	-	External Transcribed Spacer
ITS	-	Internal Transcribed Spacer
NTS	-	Non-Transcribed Spacer (Synonym = IGS)

VIII. Referenzen

- M.J. Benton, et al. (1998). 'The quality of the fossil record of vertebrates'. Pp. 269-303, in Donovan, S. K. and Paul, C. R. C. (eds), *The adequacy of the fossil record*, Fig. 2. Wiley, New York, 312 pp.
- C. M. Berteaux, et al. (2006). 'PCR and PCR-RFLP of the 5S-rRNA-NTS region and salvinin A analyses for the rapid and unequivocal determination of *Salvia divinorum*'. *Phytochemistry* 67(4):371–378.
- F. R. Blattner (1999). 'Direct amplification of the entire ITS region from poorly preserved plant material using recombinant PCR.'. *BioTechniques* 27(6):1180–1186.
- E. S. Buckler, et al. (1997). 'The Evolution of Ribosomal DNA Divergent Paralogues and Phylogenetic Implications'. *Genetics* 145(3):821–832.
- Z. H. Cai, et al. (1999). 'Molecular diversity of 5S-rRNA spacer domain in *Fritillaria* species revealed by PCR analysis'. *Planta medica* 65(4):360–364.
- M. Carles, et al. (2001). 'Chips and Qi: microcomponent-based analysis in traditional Chinese medicine' 371(2):190–194.
- S.-J. Chiou, et al. (2007). 'Authentication of Medicinal Herbs using PCR-Amplified ITS2 with Specific Primers'. *Planta Medica* 73(13):1421–1426.
- Y. Diao, et al. (2009). 'Authentication of *Panax ginseng* from its adulterants by PCR-RFLP and ARMS'. *Planta medica* 75(05):557–560.
- T. H. Eickbush & D. G. Eickbush (2007). 'Finely Orchestrated Movements: Evolution of the Ribosomal RNA Genes'. *Genetics* 175(2):477–485.
- G. Gnani, et al. (2010). 'Comparative characterization of *Santolina insularis* chemotypes by essential oil composition, 5S-rRNA-NTS sequencing and EcoRV RFLP-PCR'. *Phytochemistry* 71(8-9):930–936.
- G. M. Gongadze (2011). '5S rRNA and ribosome'. *Biochemistry (Moscow)* 76(13):1450–1464.

- D. M. Hillis & M. T. Dixon (1991). 'Ribosomal DNA: molecular evolution and phylogenetic inference'. *Quarterly Review of Biology* pp. 411–453.
- K. Lu, et al. (2005). 'Identification of *saposhinkoviae* radix in concentrated Chinese medicine preparations by nested PCR and DNA sequencing methods'. *Journal of Food and Drug Analysis* 13(3):219.
- K.-T. Lu, et al. (2007). 'Discriminating between *achyranthis bidentatae* radix and *cyathulae* radix in Chinese medicine preparations by nested PCR and DNA sequencing methods'. *Planta medica* 73(12):1322–1326.
- H. Mizukami (1995). 'Amplification and Sequence of a 5S-rRNA Gene Spacer Region from the Crude Drug *Angelica* Root'. *Biological and Pharmaceutical Bulletin* 18(9):1299–1301.
- A. S. Petrov, et al. (2014). 'Secondary Structures of rRNAs from All Three Domains of Life'. *PLoS ONE* 9(2):e88222+.
- S. Rogers & A. Bendich (1987). 'Ribosomal RNA genes in plants: variability in copy number and in the intergenic spacer' 9(5):509–520.
- P. Rubiolo, et al. (2009). 'Chemical and Biomolecular Characterization of *Artemisia umbelliformis* Lam., an Important Ingredient of the Alpine Liqueur "Genepi"'. *J. Agric. Food Chem.* 57(9):3436–3443.
- N. Sugimoto, et al. (1999). 'DNA profiling of *Acorus calamus* chemotypes differing in essential oil composition.'. *Biological & pharmaceutical bulletin* 22(5):481–485.
- K. Tamura, et al. (2013). 'MEGA6: Molecular Evolutionary Genetics Analysis version 6.0'. *Molecular Biology and Evolution* 30(12):mst197–2729.
- B. White TJ, et al. (1990). 'Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics'. *PCR Protocols: A guide to methods and applications.*(Innis MA, Gelfand DH, Sninsky JJ, White TJ, eds.). Academic Press, New York pp. 315–322.
- J. Wolters & V. A. Erdmann (1988). 'Compilation of 5S rRNA and 5S rRNA gene sequences'. *Nucleic Acids Research* 16(suppl):r1–r70.