

Molecular evolution and phylogenetic utility of non-coding DNA: applications from species to deep level questions

Thomas Borsch · Dietmar Quandt ·
Marcus Koch

Published online: 7 February 2009
© Springer-Verlag 2009

The analysis of non-coding DNA in evolutionary studies of plants is witnessing an ever-increasing popularity in addition to the widespread application of sequences of protein coding genes like plastid *rbcL* and *matK*, mitochondrial *coxI* and *matR* or the 18 and 26S parts of the nuclear ribosomal array. For the chloroplast genome, the use of universal primers annealing to highly conserved tandemly repeated tRNA genes allowed the routine amplification and sequencing of variable chloroplast spacers and the *trnL* intron since the beginning of the 1990s. Protocols for other spacers followed shortly after, facilitated by the availability of several completely sequenced green plant plastid genomes. At around the same time the use of nuclear internal transcribed spacer (ITS) sequences was proposed for

phylogenetic analyses. This has led to a large proportion of non-coding sequences available in EMBL/GenBank. For example, *trnL* intron sequences have so far been published for approximately 30% of all 450,000 species of land plants. Similar numbers can be found for nrITS sequences. Nevertheless, recent years have also witnessed a trend to look for new non-coding markers in all three genomic compartments of plants.

Organelar and nuclear genomes of plants are made up to a large extent by non-coding sequences. There is a broad spectrum of different non-coding genomic regions such as introns, spacers, satellite regions. Therefore, non-coding DNA constitutes an immense source of information which will be accessible through the more and more completely sequenced genomes. The information content of non-coding DNA has been predominantly used for phylogeny reconstruction among closely related species or genera, and in population level studies, for example to create haplotype networks or in the form of fragment size analyses of microsatellites. More recently, the application of non-coding DNA for phylogeny reconstruction at deeper levels became more prominent because well resolved and statistically highly supported angiosperm trees could be inferred based on relatively few nucleotides of plastid spacers and introns. This growing awareness for the potential of non-coding DNA markers at the same time requires a careful implementation of analytical tools in order to deal with the complexity of evolutionary patterns in non-coding sequences that are largely caused by frequent micro-structural mutations in addition to substitutions of nucleotides.

To explore the potential of non-coding genomic regions and recent developments for their analysis were the objectives of a workshop “Molecular evolution and phylogenetic utility of non-coding DNA: applications from species to deep level questions” that was held during the

T. Borsch (✉)
Botanischer Garten und Botanisches Museum Berlin-Dahlem
und Institut für Biologie/Botanik, Freie Universität Berlin,
Königin-Luise Str. 6-8, 14195 Berlin, Germany
e-mail: t.borsch@bgbm.org

D. Quandt
Nees-Institute for Biodiversity of Plants,
Plant Phylogenetics and Phylogenomics Group,
Rheinische Friedrich-Wilhelms-Universität Bonn,
Meckenheimer Allee 170, 53115 Bonn, Germany

M. Koch
Heidelberg Institute of Plant Sciences,
Division Biodiversity and Plant Systematics,
Botanical Gardens and Herbarium Heidelberg,
Faculty of Biological Sciences,
University of Heidelberg,
Im Neuenheimer Feld 345,
69120 Heidelberg, Germany
e-mail: mkoch@hip.uni-heidelberg.de

17th International Symposium on Biodiversity and Evolutionary Botany of the German Botanical Society in Bonn in September 2006. The prominent results of this workshop are published in the present volume.

Scot Kelchner deals with the specifics of models and model selection for non-coding DNA as they are required for all Bayesian and likelihood approaches to evolutionary questions. Models provide fundamental pictures on evolutionary patterns in character partitions, and in our case into the molecular evolution of genomic regions. David Morrison outlines the process of hypothesis testing in phylogenetic sequence alignment that is connected to viewing indels as mutational events that comprise multiple nucleotides at once. Helga Ochoterena then provides a parsimony perspective on homology in coding as compared to non-coding DNA sequences. This is of particular relevance in the context of aligning length variable sequences. The principle of parsimony here provides the basis for a widely used approach to assess homology. Thomas Borsch and Dietmar Quandt examine the large body of data now existing on introns and spacers of the chloroplast genome as tools for phylogeny reconstruction in land plants. Aims of their study are to evaluate the state of knowledge on finding the best performing markers with respect to resolution and support of the trees to be inferred, and to evaluate pitfalls and possibilities in widely used genomic regions. Nadja Korotkova et al. apply non-coding sequences of the *petD* group II intron to reconstructing

relationships within the angiosperm order Malpighiales, which exhibits a notoriously difficult phylogenetic problem connected to the analysis of a rapid radiation. Marcus Koch et al. investigate structural dynamics of *trnF* pseudogenes in various chloroplast genomes of cruciferous plants and discuss its application in plant evolution on species and population level. Using mosses as an example, Ute Warhmund et al. compare the phylogenetic structure in several structurally different non-coding regions of the mitochondrial genome to chloroplast loci. Mitochondrial markers in plants are considerably less understood as compared to plastid markers but especially mitochondrial introns appear to hold great promises. Miriam Calonje et al. focus on non-coding genomic regions in the nuclear genome and their potential for evolutionary analyses. Important issues are the biology of the nuclear ribosomal DNA with its consequences on the molecular evolution of the widely applied ITS region but also on the state of knowledge on other currently available non-coding markers in the nuclear genome. Thekla Pleines et al. discuss the infraspecific variability that can be found in many non-coding regions and its application to phylogeographic and speciation studies. In fact, non-coding DNA is the major source for information in any population genetic application.

We hope that the contributions from this special volume will inspire researchers to explore the full potential of non-coding DNA and to further stimulate its application in a broad spectrum of biological studies.