UNIVERSITY OF GOTHENBURG

Phylogenetic Inference and Allopolyploid Speciation

A Study of *Silene* section *Physolychnis*

Anna Petri
2012

Department of Biological and Environmental Sciences
Faculty of Science
University of Gothenburg

The thesis will be defended in public Friday September 21st at 13:00, in the lecture hall (Hörsalen) at Carl Skottsbergs gata 22B, Göteborg

Faculty opponent: Prof. Jeff Doyle, Dept. of Plant Biology, Cornell University

Examiner: Prof. Mari Källersjö, Dept. of Biological and Environmental Sciences, University of Gothenburg

ISBN: 978-91-85529-50-6

Anna Petri 2012. Phylogenetic Inference and Allopolyploid Speciation – A Study of *Silene* section *Physolychnis*. Ph.D. thesis. University of Gothenburg. ISBN: 978-91-85529-50-6

Dissertation abstract

The major theme of this thesis is allopolyploidization, with focus on *Silene* section *Physolychnis*. The evolutionary history of an allopolyploid species can be established by inferring the sister relationships of the homoeologous sequences from low copy nuclear markers. Sequence specific primers are here used to recover co-amplified genetic variants, which is an efficient alternative to the commonly used bacterial cloning.

A phylogenetic overview of *Physolychnis* is presented, and two major clades within the section are defined: the Asian/American clade and the Arctic/Siberian *S. ajanensis* group. Previously unknown ploidy levels of several taxa are inferred from sequence data, based on the number of monophyletic sequence-clusters recovered per species. Several allopolyploid taxa stemming from crosses between the two major clades are identified. Of these, *S. sachalinensis* and *S. involucrata* have indistinguishable origins, although they are geographically and morphologically distinct. Certain taxa within *Physolychnis* exhibit an extra copy of the nuclear gene *RPA2*, originating from a distantly related lineage within section *Auriculatae*. This is in strong conflict with the species relationships (inferred in several previous studies), and is best explained by introgression between lineages belonging to the two *Silene* subgenera. This finding of gene flow between such long diverged lineages may represent one of the most extreme cases presented yet.

In order to investigate whether *S. sachalinensis* and the three subspecies of *S. involucrata* originate from one or several hybridization events, transcriptome data from three distantly related *Silene* species is used to design a set of primers for 27 loci not previously sequenced in this genus. Together with five other loci, these are amplified from 43 specimens from within the parent / allopolyploid species complex, and sequenced using 454 amplicon sequencing. This method is particularly well suited for a systematic project involving allopolyploid taxa, since sequence variants are separated on the sequencing plate, and due to the possibility of multiplexing a large number of loci and individuals. Here, a two-step PCR approach is taken to attach 10 bp barcodes to the individual amplicons. This approach is time and cost efficient, but may lead to large amounts of recombinant sequences during the second PCR amplification.

Species tree inference is complicated by the presence of homoeologous gene copies within a single species. A phylogenetic tree in which identical taxon labels occur in more than one monophyletic position is here defined as a *multiply labeled tree*. Although several gene-trees-to-species-tree methods exist, the first consensus method that combines several multiply labeled gene trees into a multiply labeled genome tree is presented here. The genome tree can subsequently be folded into a species network, which describes the evolution of allopolyploid taxa.

Keywords: allopolyploidization, hybridization, next-generation sequencing, *Physolychnis*, sequence specific primers, *Silene*, species networks, transcriptomics