

# **Species Delimitation and Phylogenetic Relationships**

*A study of *Silene* sections *Atocion* and *Cryptoneuræ**

ZEYNEP AYDIN

**FACULTY OF SCIENCE**

**DEPARTMENT OF BIOLOGICAL AND ENVIRONMENTAL SCIENCES**



**UNIVERSITY OF GOTHENBURG**

Gothenburg 2014

© ZEYNEP AYDIN, 2014

**ISBN:** 978-91-85529-69-8

Printed by Kompendiet, Gothenburg-2014

*Cover Image*

Photo of a *Silene assyriaca* population occurs near Tillo, Siirt, Photographed by Zeynep Aydın, May-2009.

*To my siblings and İlgim*





# Species Delimitation and Phylogenetic Relationships

## A study of *Silene* sections *Atocion* and *Cryptoneuræ*

**Zeynep Aydın**

University of Gothenburg, Department of Biological and Environmental Sciences  
Box 461, SE-40530 Gothenburg, Sweden

### Abstract

The existence of conflicting genealogies of different genes through the evolution of species complicates the inference of phylogenetic relationships. The Multispecies Coalescent (MSC) model provides a theoretical background that account for the stochasticity in the genealogical process, thus providing systematists with a potentially objective way of testing alternative hypotheses of putative species.

This thesis focus on species delimitation under the MSC model with particular reference to *Silene* L. (Caryophyllaceae) sect. *Cryptoneuræ* Aydın & Oxelman and sect. *Atocion* Otth. A phylogenetic overview of both sections including several taxonomic conclusions are presented. Based on extensive sampling of nuclear *ITS* and chloroplast *rps16* markers across the tribe Sileneae, sect. *Atocion* and sect. *Cryptoneuræ* are shown to be distantly related, despite strong morphological similarities. Section *Cryptoneuræ* is formally described and a key to the included species is provided. Species limits within sect. *Cryptoneuræ* are evaluated with the Bayesian methods BP&P and marginal likelihood estimation (MLE) with \*BEAST using data from six putatively independent loci. MLE score comparison is found to be an efficient way to evaluate alternative hypotheses of species delimitations. The recognition of a new species, *S. ertekinii* Aydın & Oxelman is strongly supported by both approaches.

Species limits in sect. *Atocion* are investigated with the DISSECT method without conditioning on any classification defined *a priori*. MLE scores of morphological classifications estimated with \*BEAST are found to be inferior to classifications recognising strongly supported minimal clades from the DISSECT results, which reveal strong support for the recognition of several new species in the section. Two lineages which belong morphologically to *S. assyriaca* Hausskn. & Bornmüller ex Lazkov are found to be distantly related, thus being cryptic species, as no morphological and geographical differentiation can be detected. Two major, geographically structured clades are found in the section. One of the two western lineages should be named *S. atocioides* Boiss., whereas *S. aegyptiaca* (L) L.fil. belong to the eastern clade. *Silene delicatula* subsp. *pisidica* Boiss. is shown to be synonymous to *S. atocioides*. *Silene fraudatrix* Meikle, considered by current taxonomy as an endemic species on Northern Cyprus, is not clearly distinct from some mainland populations of *S. aegyptiaca* and those on Cyprus. From one of the studied loci, an ancient recombination event resulting from a hybridization event between the eastern and western clades is detected. This study is one of the first that applies the MSC model for species delimitation in plants. The strengths and weaknesses of the approach are discussed, as well as the possible consequences to taxonomy and, in the long run, biodiversity estimation.

**Keywords:** Caryophyllaceae, *Silene*, Section *Atocion*, *S. ertekinii*, *S. cryptoneura*, *S. aegyptiaca*, Systematics, Phylogenetics, Species delimitation, Multispecies coalescent, Marginal likelihood, Species tree, DISSECT.

ISBN: 978-91-85529-69-8

## LIST OF PAPERS

This thesis is based on the following papers. Papers will be referred to in text by Roman numerals as follows:

- I. Aydin Z., Ertekin A.S., Långström E. & Oxelman B. (2014). A new section of *Silene* (Caryophyllaceae) including a new species from South Anatolia, Turkey. *Phytotaxa*: in press.
- II. Aydin Z., Marcussen T., Ertekin A.S. & Oxelman B. (2014). Marginal Likelihood Estimate Comparisons to Obtain Optimal Species Delineations in *Silene* sect. *Cryptoneuræ* (Caryophyllaceae). PlosOne: in press.
- III. Aydin Z., Pfeil E.B., Jones G., Marcussen T., Ertekin A.S. & Oxelman B. (2014). Species Delineation With DISSECT: Evidence of High Species Diversity in the *Silene aegyptiaca* complex (Caryophyllaceae). Manuscript
- IV. Pfeil, E.B., Aydin, Z. & Oxelman B. (2014). Recombination provides evidence for ancient hybridisation in the *Silene aegyptiaca* complex. Manuscript

\*Paper I was re-printed with the permission from Magnolia Press.

# TABLE OF CONTENTS

## 1. Introduction

- 1.1 Species Delimitation And Species Trees
- 1.2 The Multispecies Coalescent Model
- 1.3 Multispecies Coalescent Species Concept
- 1.4 Species Tree Inference
- 1.5 Species Delimitation With DISSECT
- 1.6 Species Delimitation With Marginal Likelihood Estimates

## 2. Taxonomic Background

- 2.1 Taxonomy Of Genus *Silene* L. (*Caryophyllaceae* )
- 2.2 Study Species

## 3. Aims

## 4. Material And Methods

## 5. Results And Discussion

- 5.1 Paper I A New Section of *Silene* (*Caryophyllaceae*) Including a New Species From South Anatolia, Turkey.
- 5.2 Paper II Marginal Likelihood Estimate Comparisons to Obtain Optimal Species Delimitations in *Silene* sect. *Cryptoneuræ* (*Caryophyllaceae*).
- 5.3 Paper III Species Delimitation With DISSECT: Evidence of High Species Diversity in the *Silene aegyptiaca* Complex (*Caryophyllaceae*).
- 5.4 Paper IV Recombination Provides Evidence for Ancient Hybridisation in the *Silene aegyptiaca* Complex.

## 6. Conclusions And Future Prospects

## 7. Swedish Summary

## 8. Acknowledgments

## 9. References

## 1. INTRODUCTION

The major endeavor of Systematics is discovery and classification of the biological diversity. This requires identification and understanding of the diversification pattern to make a classification system which reflects the evolutionary history of life (Soltis and Soltis, 2003). Species are usually viewed as the basic components of such a system.

Much of biology depends on a meaningful taxonomy, correct species boundaries, and knowledge about the phylogenetic relationships among the species (Sites and Marshall, 2004; Wiens, 2007; Camargo and Sites, 2013). Nevertheless, taxonomic practice is to a large extent dependent on expert's opinion which usually primarily is based on phenotypic similarities between organisms. The lack of a formalized approach to taxonomic classification of species, as well as disagreement on the definition of the species category among taxonomists, eventually leads to widely different opinions on the numbers and limits of species.

Species should be delimited as objectively and rigorously as possible (Miralles and Vences, 2013) and perhaps for the first time in history, systematists have methods available for an objectively based taxonomy based on explicit hypotheses. Biotechnological advances have enabled the use of evidence directly from the genotype, such as DNA sequences, where ancestral-descendant character transitions can be assessed without having been affected by environmental factors. Recently, it has almost become standard practice to infer evolutionary relationships from multiple gene loci. In parallel, theoretical developments provide sophisticated tools where the observed evidence can be transformed into phylogenetic trees that not only give estimates of phylogenetic relationships, but also provide information about the historical processes that shaped the observed pattern. Nowadays, the field has advanced to the point where species divergence history can be reconstructed by tracing multiple gene genealogies that have evolved within these species (Edwards et al. 2007).

Due to these developments, species can be viewed as being statistically testable entities under explicit models, rather than being arbitrarily described by eye. The introduction of the Multispecies Coalescent (MSC) model to systematics has led to the development of species tree inference methods based on both Maximum Likelihood principles and Bayesian philosophy (Liu, 2008; Kubatko et al. 2009; Heled and Drummond, 2010). The methods provide a species tree topology together with the species divergence times estimated from multiple genes sampled from multiple individuals across a set of species. These achievements uncover tremendous new knowledge for many taxonomic groups. However they are not unproblematic. Although the power and accuracy of the methods have been documented extensively (Degnan and Rosenberg, 2009), their results are only valid given their assumptions. We assume that there is only one true evolutionary tree (or network) connecting all species on Earth and even if we could estimate such a gigantic tree with unlimited data, we would not know it definitively, as we can not directly observe history. However, we know that phylogenetic analyses on available data lets us closer to the true tree than randomly generated trees do (Daly et al. 2001).

In this thesis, I investigated species delimitations and phylogenetic relationships in the *Silene aegyptica* (L.) L. fil. and *S. cryptoneura* Stapf species groups by applying the MSC model as implemented in the Bayesian methods \*BEAST, BP&P, and DISSECT.

## 1.1 Species Delimitation and Species Trees

Species delimitation is the practice of assigning biological diversity to the species category. Accurate species delimitations are critical to many areas of biology including ecology, evolutionary biology, biogeography, and conservation biology (Pimm et al. 1995; Thomas et al. 2004, Brooks et al. 2006, Boykin et al. 2012). However, there has been considerable disagreement among taxonomists, and also by users of taxonomy on which criteria should be applied to recognize species.

Traditionally, the presence of fixed, diagnostic morphological characters has been the main information used for describing species and distinguish them from other species (Wiens and Servedio, 2000). However as such traits are affected by various environmental factors, they are interpreted differently by different taxonomists such that it results in different species taxonomies (see Sibley and Ahlquist, 1990). Species delimitation is especially problematic for recently diverged species. Incomplete reproductive isolation, little information content of DNA sequences resulting in poorly resolved gene trees and incomplete lineage sorting are problems to solve for accurate inference (O'Meara, 2009; Yang and Rannala, 2010).

Genetic data are rich sources of information concerning processes related to speciation and species delimitation. Unlike phenotypic characters, genetic variation is fully heritable. This is important for taxonomy in order to reflect ancestor-descendant relationships as envisioned by Darwin (Fujita and Leache, 2011). Genetic data are particularly important to delimit cryptic species which are indistinguishable morphologically (Isaac et al. 2004; Egge and Simons 2006; Davalos and Porzecanski, 2009; Zhang et al. 2011; Niemiller et al. 2012). Early population genetics relied upon analyses of allele frequencies from allozymes. It has recently become feasible to detect genetic variation at the DNA sequence level, which has allowed more genetic diversity within populations to be uncovered, and also exploration of historical patterns (Page and Holmes, 1998; Felsenstein, 2007). Nucleotide differences among the genes can be used to reconstruct genealogical relationships among these genes. These genealogies are known as gene trees, and represents the evolutionary histories of genes. But if the genes are sequenced from different species or populations, a common practice has been to assume that the reconstructed tree represents the evolutionary relationships of species (Takahata, 1989). Gene trees can be different from the species tree for various reasons including horizontal gene transfer, paralogy, hybridization and incomplete lineage sorting (Doyle, 1992; Maddison, 1997; Rosenberg, 2002; Maddison and Knowles, 2006, Rosenberg and Tao, 2008), so a in a more realistic model, gene trees evolve within the species tree (or network).

DNA sequences have a number of advantages in phylogenetic reconstruction, but they are not without problems. Despite biotechnological advances which enable inclusion of many genes and individuals to get resolution and accuracy of phylogenetics, the inference of species trees from sequence data has limitations. Until recently, species trees have usually been inferred by equating single estimated gene tree topologies with the species tree topology. When multiple loci have been available, the species tree topology has been inferred either via democratic vote processes (consensus methods; Jennings and Edwards, 2005), or by concatenation of the alignments of the different genes (Gadagkar et al. 2005), which relies on the assumption that all genes have evolved along the same tree. This may be reasonable for linked genes, such as different genes from the chloroplast genome, but not for unlinked genes (Wiens, 1998). Consensus methods rely on the assumption that the most common gene tree topology is equal to the species tree, but this has been shown to be

violated even under some simple scenarios (Degnan and Rosenberg, 2006).

Species delimitation and species tree inference has to be conducted in concert, because if assignment of individuals to species is erroneous, the species trees will not make sense. Recently, a number of species delimitation methods and applied methodological concepts have been developed. The combination of coalescent theory (Kingsman, 1982a,1982b) with powerful statistical inference tools have introduced a new paradigm in systematics (Rannala and Yang, 2003; Edwards, 2009) where species trees can be inferred from multiple loci. The new theoretical framework has provided the components to model the species relationships while accommodating conflicts among gene genealogies with their underlying species tree. Under this approach, species relationships of multiple populations connected by an evolutionary tree can simultaneously be estimated with their gene trees and related population parameters such as speciation times and populations sizes (Degnan and Rosenberg, 2009).

## **1.2 The Multispecies Coalescent Model**

Coalescent theory (Kingman, 1982; Hudson, 1990; Takahata, 1991) models genealogies within populations. It is developed from the realization that genealogy is usually easier to model back than forward in time (Nordborg, 2001). Given a number of individual gene copies from a single population, the coalescent traces the ancestries of gene copies back in time, until the most recent common ancestor (MRCA) of all the samples is reached (Rosenberg and Nordborg, 2002). Early applications of the coalescent model were limited to analyse genes from a single population but Rannala and Yang (2003) formulized the model to multiple populations by applying the constraint that divergence between two species can not be older than the time when they last shared alleles. This is known as the Multispecies Coalescent (MSC) model.

The integration of the MSC model to phylogenetics, has already revolutionized the field. The MSC model provides a theoretical background to reconstruct species phylogenies from a collection of gene trees by assuming any gene tree species tree discordance results from incomplete lineage sorting (Yang and Rannala 2003; Edwards et al. 2007; Carstens and Knowles 2007; Kubatko et al. 2009; Liu and Pearl, 2007). The model explores the shape and patterns of species trees by taking into account demographic parameters as population sizes and lineage divergence times. Populations in MSC model, are the ideal population of Wright and Fisher (Ewens, 1979) model with constant size, no overlapping generations, and no selection. Thus, each branch of the species tree constitute such population where gene coalescence occur randomly, while going backward in time. The MSC model also provides as a baseline to explore various causes of gene tree/species tree discrepancy. Recent developments of the model can account for variable populations size (Heled and Drummond, 2010), hybridization (Kubatko, 2009, Jones et al. 2013), gene duplication /loss (Rasmussen and Kellis, 2012), and gene flow (Hey, 2006).

## **1.3 Multispecies Coalescent Species Concept**

Species conceptualization remains as a controversial topic in systematics. Mayden (1997) listed 24 different species concepts and there are even more alternative definitions. Applications of different species concepts may result in different boundaries and therefore

different numbers of species. A unified species concept would greatly simplify a global interpretation of species (De Queiroz, 1998, 2007).

Since species delimitation requires a species concept, the increased application of the MSC model in phylogenetics naturally raises the distinction between species delimitation and species conceptualization. In the MSC model, species constitute the branches of the species tree and are in principle testable through the statistical nature of the model. Thus, the MSC model does not only provide replicable results of species delimitation but it offers a conceptual perspective to species recognition where particular species hypotheses can be tested in an objective way. Species in this model can be defined as independently evolving population lineages. This satisfies the criteria of several species concepts that all are covered by the general lineage concept (De Queiroz, 1998, 2007). Here, species are defined by no genetic exchange after the speciation event. This is similar to the biological species concept (i.e., Mayr, 1942) but in retrospect.

Although the MSC species concept involves a considerable potential to increase objectivity and stability of taxonomy, it assumes species as “ideal populations” that evolve according to model assumptions. Any speciation event among these populations is instantaneous. However actual biological populations are exposed to much more complex processes than MSC model can currently fully account for. For example, allopatric speciation is probably usually gradual (Lewis, 1966). This limitation and technical issues related to statistical inference are the major factors that hinder the MSC from being a unified species concept.

#### **1.4 Species Tree Inference**

Compared to methodology development for estimation of gene trees, methods for species tree inference is rather limited and many of those available suffer from statistical complexities (Liu et al. 2008, 2009). An ideal method would be one that statistically robust, consistent, and converge on the true species tree as more data are provided (Degnan and Rosenberg, 2009). Inference of species trees from multilocus data can be based on summary statistics or likelihood methods that differ in the extent of the information content of the data is used. Likelihood-based methods estimate the phylogeny by use of full data (Felsenstein, 2006), whereas summary statistics include methods such as “democratic vote” approaches which applies the most commonly occurring gene tree topology as the best estimate of the species tree (Liu et al. 2009). Concatenation method, which has been widely used in phylogenetics, is another summary statistic approach where multiple genes are combined in a single “super matrix” that is analyzed as single evolutionary tree (Gadagkar et al. 2005). These methods do not explicitly model the gene trees-species tree relationships.

Likelihood-based methods take full advantage of the MSC model and infer the species relationships by taking the coalescent histories into account. Maximum likelihood methods can estimate species tree by searching over species trees, computing the likelihood by summing over all possible gene genealogies for each species tree. However this is computationally very heavy (Liu et al. 2009). Currently implemented maximum likelihood methods (e.g., STEM; Kubatko et al. 2009) offer a species tree estimate based on fixed gene trees. Bayesian methods provide an estimate of species tree based on the posterior distribution inferred from prior distributions of the model parameters and the likelihood function by using the numerical method Markov Chain Monte Carlo (MCMC; Hastings, 1970). The Bayesian methods BEST (Liu, 2008) and \*BEAST (Heled and Drummond,

2010) jointly estimates species tree topology, divergence times, population sizes, and gene trees from multilocus sampled from multiple individuals across a set of species.

The likelihood-based methods summarized above, largely constitute the basis for MSC species delimitation methods. Species delimitation can also be inferred from single loci. For example, the Maximum likelihood method, general mixed Yule-coalescent (GMYC), delimits species from a single estimated gene tree by fitting within and between species branching models on the gene tree (Pons et al. 2006; Fujisawa and Barraclough, 2012). Current methods of species delimitation from multi-locus data generally fall into two classes (Ence and Carstens, 2011, Carstens et al. 2013). Discovery approaches, which are methods that do not require a priori partitioning of samples (i.e., Structurama, Huelsenbeck et al. 2011; Brownie, O’Meara 2010) before analysis. Validation approaches require samples to be assigned to a limited number of putative species (i.e., BP&P, Yang and Rannala, 2010; SpedeSTEM, Ence and Carstens, 2011; Bayesian model selection, Grummer et al. 2014; Aydin et al. 2014) prior to analysis. Validation approaches can only be used in the systems where lineages can be meaningfully defined priori, whereas discovery approaches can be applied to any system. For systems where existing evidence can not provide a clear delineation of putative lineages, the use of discovery methods is therefore necessary (Carstens et al. 2013). The newly developed species delimitation method DISSECT (Jones and Oxelman, 2014) combines discovery and validation approaches and offers a framework which uses both approaches simultaneously.

### 1.5 Species Delimitation with DISSECT

As species delimitation naturally will have a strong impact upon phylogenetic reconstruction under the MSC model, it would be advantageous to not have to define putative species *a priori*. DISSECT (Jones and Oxelman, 2014) is a newly developed method that estimates a species tree without obligate the user to define species *a priori*. Instead, DISSECT evaluates species trees as \*BEAST (Heled and Drummond, 2010) does, with the difference that every individual (or groups of individuals, that are assumed to belong to the same species) is treated as a potential species. The basic idea is that when estimated split times between those are negligible, they can be considered as belonging to the same MSC species. In this way, there is no need to restrict the space of possible species classifications to those compatible with a guide tree (as in BP&P, Yang and Rannala, 2010). In addition, the definition of parameters is unaffected, so there is no need for computationally demanding reversible model jump techniques. Following the estimation of the species tree, posterior frequencies for groups of individuals clustering below a user specified “collapsing height” value can be summarized. Any node on a species tree that is equal to or smaller than the collapsing height, will be placed into the corresponding cluster.

The DISSECT workflow can be categorized into two parts. In the first part, the method provide an estimation of the species tree, while taking uncertainty in species delimitation into account. This part works exactly as usual \*BEAST (Heled and Drummond, 2010) analysis with the exception that the usual birth-death prior for the species tree is replaced with one which results in species tree split heights with a spike density close to zero. The prior has two parameters, one which controls the collapsing height, and one which controls the number of clusters. In the second part, the method provides posterior probabilities for the clusters. This is done by the program “SpeciesDelimitationAnalyser” (see Jones and Oxelman, 2014). The classification with highest posterior probability may have low support (just like a tree topology may have low posterior probability, despite support for some individual clades may be high) and the assignments of individuals may overlap (not be



hierarchical), so therefore it is convenient to display the posterior frequencies for individuals belonging to the same cluster in a similarity matrix (Paper III, Figure 2).

## **1.6 Species Delimitation with Marginal Likelihood Estimates**

In Bayesian phylogenetics, model selection is appropriately performed via Bayes Factor comparison (Baele et al. 2012). The Bayes Factor is the ratio of the marginal likelihood of one model to the marginal likelihood of a competing model where the marginal likelihood measures the average fit of a model to the data.

The computation of the marginal likelihood of a model is a difficult computational problem, as it integrates over the parameter space (Xie et al. 2011). Until recently, estimates have been calculated using the harmonic mean of the likelihoods sampled in the posterior distribution. Although easy to estimate, as it can be directly obtained from a usual Markov Chain Monte Carlo analysis, recent studies (e.g., Xie et al. 2011) have shown that it overestimates the Marginal likelihood and fails to provide reliable results. On the other hand, two other relatively new methods, Path Sampling (Gelman & Meng, 1998, Lartillot and Phillippe, 2006) and Stepping-Stone sampling (Xie et al. 2011) have been shown (Baele et al. 2012, 2013) to outperform the harmonic mean estimator and to generate accurate results for the assessment of molecular clock and demographic models. Nevertheless these methods introduce large computational costs to the analyses. A number of recent studies (Grummer et al. 2014, Leaché et al. 2014, Aydin et al. 2014) have suggested that Path Sampling and Stepping-Stone sampling methods also can be applied to species delimitation, where each species classification is considered as a model. The difference between two such models is the number of, and allele assignments to the species tree terminal branches. In this approach, the model with the highest marginal likelihood estimate fit the data best.

## 2. TAXONOMIC BACKGROUND

### 2.1 Taxonomy of genus *Silene* L. (Caryophyllaceae)

Caryophyllaceae Juss. is a large family of 86 genera that include in 2200 species of annual or perennial herbs distributed across the globe (Bittrich, 1993). *Silene* L. is the largest genus within the family and occurs natively in temperate and alpine areas of all continents except Australia and Antarctica. The number of species included in *Silene* varies slightly according to generic delimitation but in recent accounts (Melzheimer, 1988; Greuter, 1995; Oxelman and Liden, 1995; Morton, 2005) tend to agree around 700-800. The classification by Oxelman et al. (2013) recognises nine genera within *Sileneae* with about 90% of the species classified in *Silene*.

Taxonomy of *Silene* has been highly controversial due to the homoplastic nature of many diagnostic characters (i.e., number of styles and capsule valves, calyx size, structure of the ovary and seed coat; Eggens, 2006; Oxelman and Liden, 1995; Oxelman et al. 1997; Oxelman et al. 2001). Following Otth's (1824) and Boissier's (1867) studies, an inclusive revision of *Silene* made by Rohrbach (1869) where *Silene* was divided into two subgenera; subgenus *Behenantha* (Otth) Endl. and subgenus *Silene*. Such a division is reasonably well in agreement to molecular phylogenetic studies (Oxelmann et al. 2001; Frajman et al. 2009; Rautenberg et al. 2010; Petri and Oxelman, 2011). A more recent global revision of *Silene* was presented by Chowdhuri (1957), where the genus classified into 44 sections. This classification has followed by several recent authors of Floras (e.g., Flora of Turkey and East Aegean Islands, Coode & Cullen, 1967; Flora Europaea, Chater et al. 1993; Flora Iranica, Melzheimer, 1988) with small alterations. With inclusion of many species not treated by Chowdhuri, the most recent global revision of *Silene* was presented by Lazkov (2003) where the genus organized into 43 sections and 86 series.

Using molecular sequence data, it has been shown that neither Chowdhuri's (1957) nor Lazkov's (2003) classification fit well with phylogenetic relationships (e.g., Oxelman et al. 1997; Oxelman et al. 2001; Eggens et al. 2006; Rautenberg et al. 2009; Petri and Oxelman, 2011). With the extended studies using information from several loci and including more taxa sampled, Oxelman et al. (2013) keep a dynamically updated classification of *Sileneae* which is in better agreement with the results of the phylogenetic studies.

### 2.2 Study Species

Section *Atocion* Otth was first described by Adolf Otth (1824). It included 14 species of which only a few have been considered to belong to the section by later authors. In the revision by Chowdhuri (1957), 19 diploid species of morphologically similar (i.e., sharing often glandular hairy spatulate to lanceolate leaves; compound dichasial inflorescence; petal limbs pink, entire or emarginate) annual Mediterranean taxa were assigned to sect. *Atocion*. Based on inflorescence, calyx and capsule features, Chowdhuri divided the section into three subsections, of which two have been shown to be phylogenetically distantly related (Oxelmann and Greuter, 1997; Oxelman and Lidén, 1995; Oxelman et al. 1997).

In Flora of Turkey, sect. *Atocion* (Coode and Cullen, 1967) is classified based on Chowdhuri's (1957) revision with several additions. In this study, *Silene cryptoneura* Stapf, *S. salamandra* Pamp., *S. insularis* Barbey and several other taxa that were recently assigned to section *Sedoideae* Oxelman & Greuter (Oxelmann and Greuter, 1997) or to the subgenus *Silene* (e.g., Oxelman & Lidén, 1995), are included in the section. Oxelman et al. (2013)

recognizes sect. *Atocion* as including only *S. aegyptiaca* (L.) L.fil., which is the type species, and closely allied taxa. *Silene cryptoneura* and its close relatives are classified in sect. *Cryptoneurae*. From morphological aspects, these two sections are strikingly similar. Despite their similarities, Erixon and Oxelman, 2008 found that they do not form sister groups in a chloroplast phylogeny based on c. 25 Kb of sequence alignments.

*Silene aegyptiaca* is a common plant which occurs on virgin gravelly soil in dry areas in the Eastern Mediterranean. In Flora of Turkey (Coode and Cullen, 1967), two subspecies are recognised. *Silene aegyptiaca* subsp. *ruderalis* Coode & Cullen is recognised by having an ascending, diffuse stem, and an inflated calyx, and occurs in Southeast Anatolia, Northern Syria, and Northern Iraq.

*Silene assyriaca* Hausskn. & Bornm. ex Lazkov is a novel taxon has recently been reported from Northern Iraq based on the comparison to *S. pseudoatocion* Desf. (Lazkov, 2004). It geographically and morphologically perfectly fits *S. a.* subsp. *ruderalis*.

*Silene atocioides* Boiss. is mostly found in non-cultivated gravelly habitats of South-Southwest Anatolia. Morphologically, it is highly similar to *S. aegyptiaca* and in Flora of Turkey, it is treated as synonym to *S. aegyptiaca*. However, despite their striking morphologic similarity, they are very divergent based on chloroplast DNA data (Erixon and Oxelman, 2008).

*Silene delicatula* Boiss. is a narrow South Anatolian endemic. This taxon is the only species that unambiguously can be distinguished from the rest based on morphology, with its small flowers, densely hairy leaves, and lack of petal appendages. Coode and Cullen (1967), described two subspecies for *S. delicatula* based on their indumentum features. Plant with dimorphic indumentum is recognized as *S. d.* subsp. *delicatula*, and the one with monomorphic indumentum is recognized as *S. d.* subsp. *pisidica* Coode & Cullen. *Silene d.* subsp. *pisidica* is described from Bozburun Mountain (Antalya, Gebiz) and morphologically fits well with *S. atocioides*.

*Silene fraudatrix* Meikle is known as a rare endemic restricted to a small location in Northern Cyprus (Meikle, 1977; Yıldız and Gücel, 2006; Yıldız et al. 2009). Morphological characters indicate close relationship to *S. aegyptiaca*, but it is characterized as being smaller and having a monochasial, rather than dichasial inflorescence compared to the latter taxon.

*Silene cryptoneura* is an endemic taxon occupies the virgin habitats of the medium high altitude zone of Southwest Anatolia. Although it has great morphological similarity (e.g. glandular pubescence, compound dichasial inflorescence, pink petal color, seed shape) to *S. aegyptiaca* and its close relatives, they are not forming a monophyletic group according to molecular studies (e.g., Erixon and Oxelman, 2008)

*Silene salamandra* is a rare endemic found on Rhodes Island in the Aegean Sea. Coode and Cullen (1967) described *S. salamandra* as a morphological extreme of *S. aegyptiaca* and therefore synonymized it with *S. aegyptiaca*. Carlström (1986) showed that *S. salamandra* is clearly different from *S. aegyptiaca* with its broader leaves, shorter calyx, entire petal limb and seed shape, and in fact it is very similar to *S. cryptoneura*.

*Silene insularis* Barbey occurs on the Aegean island of Karpathos. Although it is closely



**Photo 1.** Members of *Silene aegyptiaca* group. **a)** *S. aegyptiaca*, **b)** *S. aegyptiaca*, **c)** *S. assyriaca*, **d)** *S. assyriaca*, **e)** *S. aegyptiaca*, **f)** *S. aegyptiaca*. Names are in accordance with the taxonomy of Coode and Cullen (1967). Photo (c, d) by Bektaş Aydın.



related to *S. salamandra* and *S. cryptoneura*, it is characterized with smaller floral characteristics (i.e., calyx size, length of carpophore, petal limb) presumably due to autogamy (Oxelman and Greuter, 1997). *Silene insularis* is closely related to *S. salamandra* in seed structure and habit, however *S. insularis* can be distinguished from the former with smaller petals and shorter carpophore (Oxelman and Greuter, 1997).

*Silene sordida* Hub-Mor & Reese is another Southwest Anatolian endemic known from the Mugla region in Southwest Anatolia. Geographically it partly overlaps with the distribution of *S. cryptoneura*. The two taxa have previously been classified as closely related (Chowdhuri, 1957; Coode and Cullen 1967). However, ecological and some morphological features (i.e., *S. sordida* is restricted to serpentine soil, flowering later in the season, has nocturnal flowers, different seed morphology) support divergence of *S. sordida* from rest of the group.



**Photo 2.** *Silene ertekinii* Aydin & Oxelman

### 3. AIMS

The main scope of this study is to provide accurate species delimitations and to infer the phylogenetic relationships of the *S. aegyptiaca* and *S. cryptoneura* groups. For this, information from multiple gene loci are used to reconstruct species trees of both groups, primarily by using Bayesian implementations of the MSC model.

The emphasis is:

- to understand the phylogenetic relationship between the morphologically close *S. aegyptiaca* and *S. cryptoneura* groups, their level of genetic divergence and their phylogenetic position in the genus *Silene* (Paper I).
- to evaluate the species delimitations and phylogenetic relationships in the *S. cryptoneura* and closely related taxa. (Paper II).
- to evaluate the species delimitations and phylogenetic relationships in the *S. aegyptiaca* group (Paper III).
- to understand mechanisms behind strongly incongruent pattern detected in one of the genes studied in the *S. aegyptiaca* group (Paper IV).



**Photo 3.** *Silene aegyptiaca*, name applied in accordance with the taxonomy of Coode and Cullen (1967).



#### 4. MATERIAL AND METHODS

The presented study relies on DNA sequence data generated from plant material collected from wild populations during various field trips to Turkey and Rhodes, as well as herbarium samples.

The markers used in the studies are the nuclear ribosomal internal transcribed spacer (*ITS*) region (Oxelman and Lidén, 1995; paper I), the chloroplast *rps16* intron (Oxelman et al. 1997), and low-copy number nuclear regions from the nuclear RNA polymerase (NRNAP) gene family (Popp and Oxelman, 2004). The latter sequences are from intron regions of *NRPA2* and *NRPB2*, encoding second largest subunit of RNA polymerase I and II, respectively. The low-copy nuclear regions *EST04*, *EST09*, *EST14*, *EST24*, are newly developed regions from Expressed Sequence Tag (EST) libraries of *Silene uralensis* (Rupr.) Bocquet and *S. schafta* J.G.Gmel. ex Hohen (Petri et al. 2013). Primers developed from these libraries were optimized for their PCR amplification efficiency, covering six major subgroups of the genus *Silene*. DNA from two specimens of each *S. cryptoneura*, *S. aegyptiaca*, *S. nutans* L., *S. uralensis*, *S. schafta*, *S. latifolia* Poir, was used for PCR amplifications using PHUSION polymerase (Finnzymes) following the manufacturer's instructions. All reactions were run with an annealing temperature gradient ranging from 59°C to 71°C. Primer pairs producing single bands on a 1.5% agarose gel from at least four of the major *Silene* groups including *S. aegyptiaca* and *S. cryptoneura* were selected and refined. These were used for amplification of the region in rest of the specimens selected for the studies.

All the products were purified with Multiscreen PCR plates in a vacuum manifold (Millipore) and sent to Macrogen Inc. in Seoul, South Korea for Sanger sequencing. In general, sequences are obtained by direct sequencing of purified products, however in some cases the obtained chromatograms were polymorphic. Such products were sequenced with allele specific primers (Scheen et al. 2012). For RNAP regions, some sequences were obtained from cloned products. Assembly and editing of some of the RNAP sequences were done using Staden v.1.6.0 (Staden, 1996) in combination with Phred v.0.020425.c (Ewing, 1998) and Phrap (www.phrap.org). The rest of the sequences were edited using Geneious (www.geneious.com). Multiple sequence alignment was performed with MUSCLE and MAFFT as implemented in Geneious version 5.4.6 under default settings, and then manually adjusted.

Recombination events were checked using GARD (Kasakovsky et al. 2006), Dual Brothers (Minin et al. 2005), and RDP4 (Martin et al. 2010), depending on the availability of the programme.

Phylogenetic analyses were performed using Maximum parsimony, fast Maximum Likelihood, and Bayesian methods. Maximum parsimony gene trees were estimated using PAUP version 4.0b10 (Swofford, 2003). For the large data sets used in paper I, FastTree version 2.15 was used with the fastest mode with a GTR nucleotide substitution model with 20 gamma-distributed rate categories. Bayesian single gene phylogenies generated using BEAST (Drummond et al. 2006; Drummond and Rambaut, 2007). Data files were prepared in Beauti, and manually edited for implementations not covered by Beauti. Clock model, and gene tree priors were decided based on Marginal Likelihood scores were estimated via Path Sampling and Stepping-Stone Sampling methods.

Species level phylogenies were estimated using \*BEAST (Heled and Drummond, 2010). BP&P (Yang and Rannala, 2010, 2013) was used to estimate the posterior distributions of speciation events among all 15 possible rooted guide trees for the four minimal lineages in Paper II. Species limits in the *S. aegyptiaca* group were evaluated using the DISSECT (Jones and Oxelman, 2014) method, by searching over all the possible combination among individuals in the study. DISSECT analyses were conducted using Development version Beastv1.8.0, r5971 (Drummond et al. 2012). Cluster analyses were performed using SpeciesDelimitationAnalyser (Jones and Oxelman, 2014), available at [www.indriid.com](http://www.indriid.com). All Bayesian analyses were run without data to check for spurious prior distribution interactions.



## 5. RESULTS AND DISCUSSION

### 5.1 PAPER I

In this paper the phylogenetic positions of the *Silene* sections *Cryptoneurae* and *Atocion* were investigated using large numbers of *ITS* and *rps16* sequences sampled across the tribe Sileneae. The *Silene cryptoneura* group, including *S. cryptoneura*, *S. salamandra*, *S. insularis*, and *S. sordida*, was strongly supported as monophyletic and shown to be distantly related to section *Atocion* (Paper I, Figure 2), despite their morphological resemblance. Section *Atocion* in the sense of Chowdhuri (1957) and Coode and Cullen (1967) is corroborated as polyphyletic. Moreover, the phylogenetic position of the section differ between the nuclear and chloroplast phylogenies. *Silene sordida* is found to be distantly related to both the *S. cryptoneura* group and to sect. *Atocion*, and also as having different positions in the chloroplast and nuclear phylogenies (Paper I, Figure 2 and 3).

Based on the molecular and morphological evidence presented, a new section, sect. *Cryptoneurae* Aydin & Oxelman is described for *S. cryptoneura* and its closest relatives. Diagnostic characters are presented, and a key to included species is provided. *Silene ertekinii* Aydin & Oxelman, is described as a new species within the new section. In both *ITS* and *rps16* phylogenies, *S. ertekinii* is strongly supported as sister to rest of the section. Diagnostic characters of *S. ertekinii*, are compared to *S. cryptoneura* and the recently described *S. sumbuliana* Deniz & Düşen, which is considered to be a taxonomic synonym to *S. cryptoneura*.

*Silene cryptoneura* is morphologically more variable (see Paper I, Table 1) than *S. ertekinii*, which could be due to wider distribution of the former species. The two species differ most markedly in their petal limb shape and seed hilum characters. In *S. ertekinii*, the apex of the petal limb is rhomboid and it does not show any division, whereas it is more less flat and sometimes slightly emarginate in *S. cryptoneura* (Paper I, Fig. 4-6). Both species have seeds that are globose to subglobose. In *S. ertekinii*, the hilum is almost rounded with no differentiation at the margins, whereas it is more less rectangular with two twisted sides in *S. cryptoneura* (Paper I, Figures 3B, 4B).

### 5.2 PAPER II

By using data from five low-copy nuclear genes, and the chloroplast *rps16* intron, a Bayes factor approach (Grummer et al. 2014) was used where a range of possible classification models were compared based on their marginal likelihood scores. The performance of different marginal likelihood estimation methods (path sampling, stepping stone and harmonic mean) and the Akaike Information Criterion (AIC) were explored. Using \*BEAST, nine different classification models were evaluated. For each model marginal likelihood scores were estimated. The path sampling and stepping-stone sampling methods strongly supported models separating *S. ertekinii* from the rest. All the 15 possible guide tree topologies of the four minimal lineages were evaluated with BP&P (Rannala and Yang 2013). The recognition of *S. ertekinii* was strongly supported also by these analyses.

Bayesian methods have been advocated as being more objective compared to traditional taxonomic applications of species delimitation (Fujita and Leaché, 2011, Fujita et al. 2012) and they are being increasingly popular. Several recent studies (Kubatko et al. 2011;

Harrington and Near, 2012; Satler et al. 2013; Camargo et al. 2012) have applied a number of different methods including Bayesian methods to infer species limits in various taxonomic groups. Carstens et al. (2013) have discussed that species limits should be evaluated by using a wide range of available methods and decisions should be made by trusting on observable congruence across methods, as this will give robustness to a particular species classification. However, results from each method are only valid under its own assumptions. On the other hand, use of many different methods raise the difficulty of interpreting results, especially when there is large incongruence among these. Therefore, if an estimate of a species phylogeny is the goal, species should be delimited to maximize the fit to the particular phylogeny model. Marginal likelihood estimates for alternative species delimitation models under the MSC can be compared to identify the optimal species classification for the group under study (Grummer et al. 2014; Leaché et al. 2014).

Similar to Grummer et al. (2014), we employed marginal likelihood estimation as used in formal model selection (e.g., Baele et al. 2012), to compare different classification models implemented in \*BEAST. Marginal-likelihood scores estimated for each species delimitation can vary depending on the estimator used to calculate them. The stepping stone and path sampling methods gave strong support for the recognition of the eastern samples as a distinct species (Paper II, Figure 3). Marginal likelihood estimates calculated by the harmonic mean method contradicted the results of the stepping stone and path sampling methods. The AICM results reminded of those from harmonic mean but had higher variance. Baele et al. (2012) argued that one should use the stepping stone and path sampling methods, and avoid harmonic mean and AICM. Baele et al. (2013) also stated that it is important that analyses are performed with proper priors (integrating to 1). On the other hand, it was shown that (Baele et al. 2013b) the accuracy of marginal likelihood estimates increases if one uses a stepping stone approach to create a path between the two competing models, compared to marginal likelihood estimation of individual models, but at a significant extra cost in terms of computational demands. The results of Grummer et al. (2014) show that the approach used by us (paper II, paper III) is valid at least in some situations, but more studies applying the “Marginal likelihood estimate” approach on species delimitation would be beneficial.

Arrangement of the guide tree has critical importance for BP&P outcomes (Leaché and Fujita, 2010). When alleles can be assigned to putative species unambiguously, applying a species tree estimation method can serve as selection procedure for choosing the guide tree. However, this also requires the guide tree to be estimated correctly, which may be hard because of poor information content of the terminals. The marginal likelihood estimate method does not rely on a fixed tree topology and alternative delimitation models do not have to be nested. A potential problem in our comparisons is that the \*BEAST model is implemented only for two or more species (Heled and Drummond, 2010) so the comparison with the one-species classification may be affected by other model differences. Grummer et al. (2014) used an outgroup species to overcome this problem. In our case, the genetic distance to any other species are large (Oxelman and Liden, 1995), so other problems pertaining to difficulties in reconstructing clocklike trees with long branches may be introduced if such an outgroup is included.

The results from our study show some support (Paper II, Figure 3B, C) for *S. cryptoneura* being distinct from the *S. ertekinii* and the island lineages. The poor resolution for the position of the island lineages may be due to poor sampling, which makes it difficult to clearly resolve the phylogenetic position of these three species in the group. In particular, *S.*

*cryptoneura* and *S. salamandra* are very similar morphologically, whereas *S. insularis* is easily recognized by its smaller floral parts. Despite the similar morphology and habitat requirements, the observed genetic differentiation between *S. cryptoneura* and *S. ertekinii* suggests that the Bey Mountain range has acted as a geographic barrier against gene flow or hybridization (Aydin et al, 2014). In agreement with the current taxonomic recognition of *S. salamandra* and *S. insularis* (Oxelman et al. 1997) the island species turned out as sister lineages sharing a common ancestor with *S. cryptoneura*, although the support for this relationship was poor.

Paper II provides support for the recognition of the newly erected (paper I) species *S. ertekinii*. It also concurs with (Grummer et al. 2014) in that marginal likelihood estimation of different species delimitation models may provide an important source of information to taxonomy, and be a valuable validation approach for choosing among species classifications when attempting to reconstruct phylogenies under the MSC model.

### 5.3 PAPER III

Species delimitation and phylogenetic relationships in sect. *Atocion*, sensu Oxelman et al. (2013), were investigated using the Bayesian methods \*BEAST and DISSECT. These methods were then combined with the marginal likelihood approach to compare alternative classification models and identify the optimal species delimitations in the group.

Three different steps were followed. First, the marginal likelihoods of three different morphology-based classifications of sect. *Atocion* were estimated with \*BEAST using path sampling and stepping-stone sampling. Second, all the individuals in the study were analysed by DISSECT without specifying any prior classification knowledge. Third, a classification model compatible with the DISSECT results was analysed with \*BEAST, and found to have much higher marginal likelihood scores than the models based on the morphology-based classifications.

Without conditioning on *a priori* classifications, DISSECT is a useful tool for species delimitation, given that model assumptions are fulfilled. Our results demonstrate the strong impact of prior conceptions about species limits on the estimated phylogeny, as have been suggested by several previous authors (Zhang et al. 2011; Camargo et al. 2012b; Edwards and Knowles, 2014). The phylogenetic trees obtained from DISSECT (Paper III, Figure 2) and \*BEAST with species delimitations compatible with the DISSECT results (Paper III, Figure 4a) are widely different from those based on previous taxonomic classification based on morphology (Paper III, Figure 3).

We evaluated 75 individuals sampled across the morphological and geographical variation of the *Silene aegyptiaca* group, with each individual as a potential species by searching over all possible combinations among them. The basic underlying idea is that closely related individuals will form shallow clusters of such "species", where the heights of the splits are small enough to be negligible (Jones and Oxelman, 2014). The results indicated a large number of such clusters in our data. Despite using a beta prior distribution with a peak density around 4 clusters, the posterior was much larger (paper III, Figure 2), indicating strong signal in the data favoring many more species than current taxonomy recognises.

The MSC model assumes random mating among individuals, and instantaneous speciation (i.e., no migration is allowed after species split). Recent simulation studies have shown that topology estimates of MSC implementations (e.g., \*BEAST) are robust to migration

between sister species, but not to time and population size estimates (Heled et al. 2013; Leaché et al. 2014). If gene exchange occur between non-sister species, also the topology estimates will be grossly inaccurate. The impact of violations of the assumption of random mating within species has to our knowledge not been studied extensively. If speciation in the *S. aegyptiaca* group is gradual and allopatric, then we should perhaps expect that individuals sampled in geographical proximity should tend to cluster together more frequently in the DISSECT analysis. This is also indeed what we often observe. Under a gradual speciation model, we would expect less traces of hybridization between diverging populations ("species") as we move towards the root in the species tree. Our results seem compatible with such a scenario. We have some clusters of individuals that have support as belonging to the same MSC species. Often, there is not much hierarchical structure of these clusters, as evidenced by the low posterior probabilities of clades until a certain point (paper III, Figure 2), where we were able to identify six strongly supported clades, which we subsequently classified as species (together with the remaining three singletons) in our classification IV. Thus, there seem to be no evidence in our data for hybridisation between these nine units.

The species tree estimated from DISSECT (Paper III, Figure 2) and the \*BEAST species tree based on classification IV (Paper III, Figure 4a) differed in that the strongly supported lilac and yellow clade (Paper III Figure 2 ) in the DISSECT tree is not present in the \*BEAST tree, as the blue clade is nested within. Under the MSC assumptions, DISSECT and \*BEAST are expected to give the same (or very similar) results (Jones and Oxelman 2014), if species assignment in \*BEAST is correct. The only difference between the methods is that DISSECT uses a special birth-death prior for the species tree, which has two extra parameters (collapseHeight, defines the height under which there is a high prior density and collapseWeight, which influences the number of clusters, defined by having split heights less than the collapseHeight).

The marginal likelihood estimates (MLE) for the three morphological classification models clearly favored model III (paper III, Figure 3c), which recognizes *S. atocioides* and *S. assyriaca* (= *S. aegyptiaca* subsp. *ruderalis*). This model contradict the proposed subspecies division in *S. aegyptiaca* and *S. delicatula*, but favor more species in the *S. aegyptiaca* group. However the DISSECT-compatible classifications are favored over this classification with more than 300 units of MLE scores (paper III, Figure 5).

MLE comparison is useful for comparing the species delimitation models but impractical if one wants to explore the entire space of classifications, because only a limited number of models can be compared with reasonable computational efforts. Our approach was to use DISSECT to explore the species tree space while taking uncertainties in species assignment into account, and then define classifications compatible with the DISSECT results which can be compared with existing taxonomic classifications and/or classifications based on other criteria. With perfect data, i.e., no model violations, informative data, and convergence of the MCMC runs, one could be confident to find the best classifications, but still, it might be useful to quantify the magnitude of differences. Kass and Raftery (1995) devised levels for this, and in our case, the DISSECT classifications heavily outperformed the traditional, morphology-based classifications. We propose that this approach is superior to the one suggested by Yang and Rannala (2010), where individuals are first classified into "minimal" clusters based on the available evidence or some genetic threshold. Olave et al. (2014) using simulated data, reported high sensitivity of this latter approach to errors, especially in the first steps.

It is not immediately obvious what is the preferable course of action taxonomically for the *S. aegyptiaca* group based on the results in papers III and IV. Recognizing all clusters as taxonomic species would result in very narrow clusters, that are not separated well from other such clusters (Paper III, Figure 2). A more conservative approach could be to apply the criteria we used (recognize strongly supported clades of individuals as species) to get classification model IV. This was rather straightforward in our case, but it may not be the case in other studies.

The presence of two highly divergent lineages within the morphologically and geographically coherent *S. assyriaca* (= *S. aegyptiaca* subsp. *ruderalis*) was one of the most remarkable results of our study. The DISSECT analyses showed relatively strong support for the occurrence of two phylogenetically widely separated clusters (paper III, Figure 2) that have no clear morphological nor geographical differentiation. To our knowledge, they are indistinguishable except at the DNA level, and thus an example of "true" cryptic species. *Silene aegyptiaca* subsp. *ruderalis* and *S. assyriaca* are heterotypic synonyms, so their type specimens need to be studied for proper name designation.

The narrow endemic *S. delicatula* is the most distinct species of the *S. aegyptiaca* group from a morphological point of view, having considerably smaller flowers adapted to autogamy, and also lacking the peculiar petal appendages otherwise characteristic for the group. The DISSECT analyses revealed it as nested in the eastern clade albeit with poor support. Nevertheless, under a traditional species classification (Paper III, Figure 3b, 3c) it would erroneously be interpreted as sister species to the rest.

The Cyprian endemic *S. fraudatrix* was shown as poorly distinct from the group which includes some Central Anatolian and Middle East populations and other populations from Cyprus. This species is known restricted to a small location in north of Cyprus (Yıldız et al. 2009), and morphologically distinct (Meikle, 1969). Our analyses strongly supported *S. fraudatrix* as belonging to and perhaps nested within the main Eastern group (yellow lineage in paper III, Figure 2). If this geographical pattern is consistent, the name should probably be *S. aegyptiaca*, as the neotype of *S. aegyptiaca* (Cafferty and Jarvis 2004) was collected in Jerusalem area where we also have two samples (3505 and 6089) from. These two samples from Jerusalem were shown as belong to the yellow lineage (in paper III, Figure 2) which also includes the other samples from Middle East and most of the samples from central-south Anatolia.

#### 5.4 PAPER IV

A recombination event in the EST09 locus is inferred to have taken place between the two major lineages of *Silene* sect. *Atocion*. This is supported by several different (i.e., RDP, MaxChi, Chimaera) recombination detection methods. Phylogenetic analysis of the partitions on either side of the inferred breakpoint produced supported conflicting topologies consistent with the inferred recombination (Paper IV, Figs. 2 and 3). The divergence times strongly support an ancient hybridization event over recombination between paralogs or deeply coalescing alleles.

We found clear evidence for ancient recombination in the EST09 gene between the two main clades in the *Silene aegyptiaca* group. These analyses implicate the pink group (paper IV, Figs. 2 and 3) as the closest extant relative of the donor lineage of the 3' partition of EST09. The six genes analysed in paper III place the individuals carrying the recombined EST09 sequences in a background consistent with other members of the yellow group (Paper IV, Figure 1). This suggests that a proportion of the genome has successfully introgressed from the pink group into the yellow group.

The monophyly of the three sequences in the 3' part of EST09 suggests that a single successful introgression event took place, with subsequent divergence of allele lineages after the event. The lack of monophyly of these sequences in the 5' part of the gene is compatible with this scenario, if the 3' donor sequence was independently recombined twice (to explain the two supported different positions in Fig. 2) with different alleles present in the host species (yellow) after the introgression event. Recombination hot spots are well known (Lichten and Goldman, 1995; Hey, 2004), so recombination in approximately the same position is not unexpected.

The timing of the recombination event is consistent with ancient hybridisation that has brought together disparate alleles into the same genome and allowed meiotic recombination to occur. Alternative explanations, either of paralogous or deeply coalescing alleles as the source of the recombined sequences, both require several additional *ad hoc* assumptions. For paralogy, a gene duplication (one event), loss of copies to leave a single copy per individual (several events), and finally two unequal recombinations between the paralogous copies (two events) would be required to explain the observations. A similar number of events is also required for recombination involving a deeply coalescing allele.

Furthermore, the timing we observe is not consistent with paralogy or deeply coalescing alleles as the source, because the divergence between the source lineage and the recombinant partition is around the same age as the divergence between groups (yellow) and (lilac), rather than earlier (i.e., earlier than speciation of all species here) as expected for a paralogue or deeply coalescing allele drawn from group (yellow) that has its closest relationship to group (pink).

These explanations are less parsimonious than introgressive hybridisation, which requires the introgression itself (one event), and two ordinary meiotic recombinations (two events). The hybridisation is also compatible with the geographic proximity of group (pink) and group (yellow), with the individuals bearing the recombined sequences found along the western-most part of the distribution of group (yellow), closest to group (pink) individuals (Paper IV, Figure 5). We also rule out template switching during amplification as a viable explanation of the origin of the recombined sequences. The most likely source of an alternative template, a paralogous locus, generates the same *ad hoc* assumptions as discussed above, and is rejected here.

Contamination with other samples in this study can also be ruled out, because the branch lengths of the recombined piece in the 3' partition analysis shows many changes since recombination, which is clearly at odds with this explanation. Contamination should result in identical or near identical sequences in the recombined part to the source of the contamination (which is most likely from the same study, given PCR primer specificities).

## 6. CONCLUSIONS AND FUTURE PROSPECTS

Despite recent developments, evaluation of species boundaries with molecular data is still complicated. The results must be interpreted according to the assumptions of the model which are critical to the outcomes of the study. The MSC model species unit are the branches of the species tree, which in turn are single, ideal Wright-Fischer populations. Nevertheless, the complexity of actual biological processes is much larger than the MSC model can account for, thus violations to the model may affect the inferred species history in particular estimates of the model parameters. Reid et al. (2013) discussed the rare examination of the absolute data fit to the MSC model and showed that many empirical data sets poorly fit to the MSC model, something which potentially can generate inaccurate phylogenetic estimates. The low posterior probabilities for the shallow branching events in Paper III, Figure 2 may be due to violation of the assumptions of the MSC model. As DISSECT treats single individual as species, the model will be violated if the evolutionary history of these individuals is not strictly hierarchical. This will be the case any time that two individuals or more actually belong to the same species. However, in such situations, the branching heights will be so small that the violations probably have mild or negligible effects to the species tree below. If, as the results presented in this thesis indicate, speciation is gradual, the effects are less clear. Heled et al. (2013) and Leaché et al. (2014) have recently performed simulations to study the impact of migration on MSC implementations. In short, topology estimates are robust to migration between sister species, but estimates of divergence times and population sizes are biased. When migration occurs between non-sister species, also topology estimates become distorted. Misspecification of species assignments are equivalent to migration events. Although it has not yet been studied directly, it seems likely that DISSECT will produce more accurate species trees than regular \*BEAST runs, if the species assignments are poor.

In comparison to the conceptualization of species, systematists have paid relatively little attention to delimiting species using formalized methods (Wien and Servedio, 2000; Sites & Marshall, 2004). As a consequence there is usually conflict on which methodology or data should be used to recognize species (De Queiroz, 2007). With the recent advances in biotechnology, genetic data enable us to see the history of gene flow, and therefore conclude objectively about species boundaries. On the other hand, if speciation is a continuous process, we will necessarily sometimes observe cases with ongoing, "incomplete" speciation.

With the awareness that trees from different parts of the genome have conflicting branching patterns, our studies contribute to the emerging standard to apply multilocus, and even whole genome (Boussau et al., 2013; Marcussen et al., 2014) information to phylogenetic inference (Degnan and Rosenberg, 2006; Liu et al. 2008; Rosenberg et al. 2008). Using multiple gene sequences, a species tree for a group of closely related organisms can be estimated under the MSC model, with prior knowledge and uncertainty of included parameters incorporated in the result. The MSC model offers a powerful statistical framework, where the species are the branches of the species tree. This enables objective testing of alternative hypotheses of species delimitation. The DISSECT method eliminates the requirement of *a priori* assignment of individuals to putative species and therefore potentially can increase the accuracy of species delimitation and, consequently, tree analyses. Nevertheless, several aspects need to be improved. Among these are the need to account for various biological processes that cannot easily be ignored for many groups, such as gradual speciation and hybridization. The marginal likelihood estimate approach is an efficient way to select among

a limited number of alternative species delimitations. The DISSECT method seems to be an important improvement to the \*BEAST method, as it can take uncertainties in species delimitations into account.

The results presented here enabled us to recognize previously unrecognized distinct lineages in *Silene* sections *Atocion* and *Cryptoneuræ*. We formally described a new species, *S. ertekinii*, and showed that its recognition is supported by the MSC model. We also showed that morphological classification of sect. *Atocion* is not congruent with phylogenetic relationships inferred from our data. The number of MSC species is underestimated, but formal reclassification of the entire group will require DNA sequencing of type specimens to be accurate. Even so, it is questionable whether a "stable" classification will be attained, and if there at any given point in time, will be a fixed number of entities we should call species.



## 8. SWEDISH SUMMARY

Vad är en art? Trots att begreppet är relativt alldagligt och av central betydelse inom biologi finns ingen entydig definition. På senare år har det utvecklats en populationsgenetisk modell (koalescent-teori) som matematiskt beskriver hur genträäd, dvs hur olika alleler är besläktade med varandra i ett fylogenetiskt träd, fördelar sig inom en population. I sin ursprungliga version görs mycket enkla antaganden som att fortplantningen är slumpmässig (alleler "väljer" sina föräldrar slumpmässigt), generationerna överlappar inte och migration förekommer inte. Även om dessa antaganden kan tyckas orealistiska så har modellen visat sig mycket användbar inom populationsgenetiken. Man kan även lägga till parametrar i den matematiska modellen för att den skall bli mera realistisk. I "Multispecies Coalescent" (MSC)-modellen sätts flera populationer, vi kan välja att kalla dem arter, ihop i ett arträd. Koalescent-modellen förutsäger att olika genträäd kommer att se olika ut, om vi till exempel väljer en allel per art. MSC-modellen gör att vi kan rekonstruera arternas släktskap till ett arträd. Man kan säga att genträden utgör data för att rekonstruera arträd. För att detta skall låta sig göras måste man dock veta vilken art de olika individerna hör till, annars kan resultaten bli missvisande.

I denna avhandling används framför allt genetiska data i form av DNA-sekvenser från olika gener för att studera artavgränsningar i ett par artgrupper av glimsläktet (*Silene*), som förekommer i östra Medelhavsområdet (Grekland, Turkiet och länder i Mellanöstern som gränsar till Medelhavet). I ett av de underliggande arbetena jämförs sekvenser från två olika gener, en från kloroplastgenomet, samt ITS-regionen från kärngenomet med ett stort antal (6-700) sekvenser från många andra arter i glimsläktet. Trots att arterna i de två artgrupperna morfologiskt är mycket lika, visar resultaten på ett övertygande sätt att de inte är varandras närmaste släktingar. För den ena gruppen fanns ett taxonomiskt namn tillgängligt, nämligen sektion *Atocion* Otth. En sektion är en rang mellan art och släkte i den taxonomiska hierarkin, och sektionen *Atocion* beskrevs först av Adolf Otth redan 1824. Olika systematiker har dock inkluderat olika arter i sektionen, och baserat på resultaten här visar det sig mest lämpligt att enbart inkludera typarten *Silene aegyptiaca* och dess närmaste släktingar. En typart är det som avgör vilket namn som skall användas på en sektion, det säger dock inget om vilka andra arter som skall inkluderas. För den andra gruppen fanns inget tillgängligt sektionsnamn, så den nybeskrevs därför formellt som section *Cryptoneurae* Aydin & Oxelman, med *Silene cryptoneura* som typart.

Sektionen *Cryptoneurae* studerades närmare med avseende på artavgränsning i ett annat arbete. Sex olika regioner sekvenserades för ett antal individer som traditionell taxonomi klassificerat som tre arter, nämligen *Silene insularis*, som bara finns på den grekiska ön Karpathos, *S. salamandra* från Rodos, och *S. cryptoneura* från sydvästra Turkiet. Våra egna observationer från studier av växterna i fält och i herbariesamlingar hade dock redan lett oss till att beskriva den nya arten *Silene ertekinii* Aydin & Oxelman, som endast förekommer öster om berget Bey i Lycien. Vi använde oss av två olika implementeringar av MSC-modellen, för att avgöra om vårt beslut var överensstämmande med denna.

I den ena, BP&P, bestämmer man sig först för ett guideträd, och låter sedan en så kallad "reversible jump Monte Carlo Markov Chain" skatta sannolikheten för de olika artklassificeringar som är kompatibla med guideträdet. Monte Carlo Markov Chain (MCMC) är en metod som används för att skatta så kallade Bayesianska sannolikheter. Dessa bestäms av data och sannolikhetsmodell, men också av vad man tror sig veta om sannolikheterna på förhand. Detta är en gammal idé, men det är mycket svårt att beräkna

sådana sannolikheter exakt. MCMC-metoden är dock en kraftfull metod för sådana skattningar. En sannolikhetsmodell består av ett antal parametrar. I fylogenetisk analys är grenar, vars längder skall skattas, exempel på sådana parametrar. Eftersom antalet grenar i arträdets förändras om man till exempel slår samman två systergrenar i guideträdet, fungerar inte MCMC-metoden, och man måste lägga till "reversible model jumps". Detta fungerar fint så länge de olika modellerna är nestade i varandra, vilket de är i guideträdet, men inte annars.

Den andra metoden vi använde var att helt enkelt skatta arträdets med hjälp av en implementering av MSC-modellen som heter \*BEAST. För att få ett mått av hur bra data passar till den modell man valt, kan man skatta integralen av sannolikheten för modellen utifrån alla möjliga parametervärden. Detta är mycket svårt, och man har nyligen visat att en del populära metoder är otillförlitliga. Andra, bättre metoder har dock utvecklats. Fördelen med dessa beräkningar är att man kan jämföra olika modeller utan att parametrarna behöver vara hierarkiskt besläktade. Vi applicerade dessa metoder på sektionen *Cryptoneuræe*, och kunde på så sätt jämföra klassificeringar som inte kunde jämföras med BP&P. Bägge metoderna gav dock starkt stöd för *Silene ertekinii* som en egen art.

Inom sektionen *Atocion* använde vi sekvensdata från sju regioner och 75 individer insamlade från hela det geografiska utbredningsområdet för gruppen, och dess morfologiska variation. I sektionen *Atocion* har mellan tre och fem arter urskiljts morfologiskt. Det vore dock en klar fördel om man kunde komma fram till en optimal klassificering utan att vara begränsad till ett fåtal på förhand bestämda. En alldeles ny metod, DISSECT, bygger på idén att man låter MCMC söka igenom rymden med alla möjliga arträd, precis som \*BEAST gör, men man definierar varje individ som en egen art. Om de skattade förgreningarna har grundare höjd än ett visst värde, ser man dem som matematiskt försumbara. Individer som grupperar sig under en sådan höjd kan då sägas tillhöra samma MSC-art. Analysen visar att antalet sådana "arter" är mycket stort, kanske ett fyrtiotal, och att det finns stora konflikter i arträdets topologi hur de är besläktade. Dock vid en viss höjd, finns det grupper av individer som har mycket starkt fylogenetiskt stöd (och genträden är starkt överensstämmande). Genom att klassificera dessa grupper som 9-10 arter och jämföra denna klassificering med de morfologiska finner man att den DISSECT-deriverade klassificeringen passar data och modell mycket bättre.

Våra resultat visar att det som morfologiskt klassificeras som *S. assyriaca* Hausskn. & Bornmüller ex Lazkov, i själva verket är riktiga kryptiska arter, eftersom ingen morfologisk och geografisk differentiering kan detekteras. Två stora, geografiskt korrelerade klader kunde identifieras. En av de två västra linjerna kan lämpligen namnges *S. atocioides* Boiss. *Silene aegyptiaca* (L.) L. tillhör den östra kladen. *Silene delicatula* subsp. *pisidica* Boiss. visades vara synonymt med *S. atocioides*. *Silene fraudatrix* Meikle, anses av nuvarande taxonomi som en endemisk art på norra Cypern men skiljer sig dåligt från cypriska och vissa fastlandspopulationer av *S. aegyptiaca*.

Från en av de studerade DNA-regionerna kunde en gammal rekombinationshändelse till följd av en hybridiseringshändelse mellan de östra och västra kladerna upptäckas. Denna upptäckt, tillsammans med den dåliga fylogenetiska strukturen bland de små MSC-arterna detekterade av DISSECT, kan förklaras av en artbildningsmodell där arterna skiljs åt gradvis. MSC-modellen antar att arterna skiljs abrupt, efter förgreningen förekommer inte något utbyte. Detta är en svaghet hos MSC-modellen som gör att man inte kan förvänta sig otvetydiga svar, ens med väldigt mycket data. Trots detta utgör den en viktig och användbar

modell för att förstå de evolutionära släktskapsförhållandena mellan närbesläktade grupper. Denna studie är den, så vitt känt, första någonsin som använder MSC modellen för artavgränsning hos växter. Med tanke på hur lätt det numera är att få fram många jämförbara DNA-sekvenser från många individer kommer sannolikt denna modell och förfiningar av den få mycket stor betydelse för framtida taxonomi och skattningar av biologisk mångfald.

## 8. ACKNOWLEDGEMENTS

First of all, my deepest thanks to my supervisor Bengt Oxelman, whose excellent knowledge, and supervision made this work possible. He provided the opportunities, taught me all I know about phylogenetics. He encouraged me to take steps forward, was patient and supportive, and never lost his belief in me, even in the moments I have little belief in myself.

Many thanks to my co-supervisors, A. Selçuk Ertekin & Bernard E. Pfeil, who made great contributions, made everything easier and solved the problems on the way. Field work would have never been that enjoyable without A. S. Ertekin, and I would have never seen how elegant the approaches to phylogenetic puzzles can be if I didn't meet Bernard.

My special thanks to Mariette Manktelow, who advised me to come to Sweden, and is being my first contact with Bengt.

A big thanks to;

Thomas Marcussen; for all the contributions, ideas, discussion, especially the fast feedback and comments on the manuscripts, and the company of great times in Gothenburg. Vivian Aldén; queen of the labs, my sweet office mate of most times, Graham Jones; for theoretical advises and help on the DISSECT analyses, Mari Kallersjö; for all the support and keeping things on track, Ingela Dahllöf; for the generous support and facilities at Botanhuset.

People at University of Dicle: Murat Kızıl, who has been very inspiring and supportive, guiding me for study possibilities abroad, and teaching me to dare go after my goals. Ahmet Onay, Hasan Çetin Özen, and Murat Biricik for sharing their experience, and giving me motivation.

...It was a tough work but my Botan team was just as tough. Thank you all for sharing the burden, I'm grateful to you all. So, the forward players:

Anna; who has a much bigger heart than she shows, my handbook on "Swedish life", organizer of wonderful moments usually including great trips. Filipe; who can be described with only one word, mine would be "gentleman", a really real one. Mats; who (I know you wouldn't believe) has magic power, well just ask him "the most impossible thing you can think of that the cluster could do for you" See!. Tine; who is a combination of Norwegian beauty and kindness, the boss of cloning. Yann; who I was disturbing only three times a day with my instantaneous questions (often related with figure editing) but never stopped smiling at me (and correcting my English). Ylva; who is an angel from "lands of angels" ...

In the midfield: Henrik N; for the cheery office company, I ensure you that there has not been a single day I didn't laugh when he was around. Elisabeth S; for the best laugh ever. Systematik group; Alex A., Alex Z., Atefeh, Bente, Cajsa, Christine, Claes G., Claes P., Climbié, Elisabeth G., Ellen, Hamid, Ivana, Jonna, Maria, Narjara, Roger, Ruud for keeping the work environment that friendly.

And the defense: Azeez, Andrei, Cornelia, Daniela, Elham, Henrik A., Jila, Kemal, Lan, Mohamed, Panos, Patrick, Sazzad, Somnath, Thomas H., Triranta and all Botan people ...

A big thanks to Sven, who has been the quick remedy of any conflicts I had in Botan, Ingela L. Carol, Linn, Olga and Niklas for taking care of various issues.

Outside Botan,

Jeanette, my dear friend, miracle in the kitchen, I can't find words to tell the pleasure I have for meeting you and Stina, David, Uve, and little ones.

My mother, all my siblings and my daughter, the most precious of my life, thank you for all the care and endless support. Bektaş Aydın, for his help and nice photos, Semsettin Turga for the long friendship we came along, and all the friends in Diyarbakır.

## 9. REFERENCES

- Aydin Z, Ertekin AS, Långström E, Oxelman B. 2014. A new section of *Silene* (Caryophyllaceae) including a new species from South Anatolia, Turkey. *Phytotaxa*, in press.
- Aydin Z, Marcussen T, Ertekin AS, Oxelman B. 2014. Marginal Likelihood Estimate Comparisons to Obtain Optimal Species Delineations in *Silene* sect. *Cryptoneuræ* (Caryophyllaceae). *PlosOne*, in press.
- Baele G, Lemey P, Bedford T, Rambaut A, Suchard MA, Alekseyenko AV. 2012. Improving the accuracy of demographic and molecular clock model comparison while accommodating phylogenetic uncertainty. *Molecular Biology and Evolution*, 29:2157-2167.
- Baele G, Li WLS, Drummond AJ, Suchard MA, Lemey P. 2013. Accurate Model Selection of Relaxed Molecular Clocks in Bayesian Phylogenetics. *Molecular Biology and Evolution*, 30:239-243.
- Baele G, Lemey P, Vansteelandt S. 2013. Make the most of your samples: Bayes factor estimators for high-dimensional models of sequence evolution. *BMC Bioinformatics*, 14:85 doi:10.1186/1471-2105-14-85
- Bittrich V. 1993. Caryophyllaceae. Magnoliid, Hamamelid, and Caryophyllid families. The families and genera of vascular plants, In K. Kubitzki, J. Rohwer, and V. Bittrich [eds.], 2:206–236.
- Boissier E. 1867. *Flora Orientalis sive enumeratio plantarum in Oriente a Graecia et Aegypto ad Indiae fines hucusque observatarum*. Genevae et Basileae.
- Boussau B, Szöllosi GJ, Duret L, Gouy M, Tannier E, Daubin V. 2013. Genome-scale coestimation of species and gene trees. *Genome Research*, 23:323-330.
- Boykin LM, Armstrong KF, Kubatko L, De Barro P. 2012. Species delimitation and global biosecurity. *Evolutionary Bioinformatics online*, 8:1.
- Brooks TM, Mittermeier RA, da Fonseca GAB, Gerlach J, Hoffmann M, Lamoreux JF, G. MC, Pilgrim JD, Rodrigues ASL. 2006. Global Biodiversity Conservation Priorities. *Science*, 313:58-61.
- Cafferty S, Jarvis Charles E. 2004. Typification of Linnaean plant names in *Caryophyllaceae*. *Taxon*, 53:1049-1054.
- Camargo A, Morando M, Avila LJ, Sites JW. 2012. Species delimitation with ABC and other coalescent-based methods: a test of accuracy with simulations and an empirical example with lizards of the *Liolaemus darwini* complex (Squamata: Liolaemidae). *Evolution*, 66:2834-2849.
- Camargo A, Avila LJ, Morando M, Sites JW. 2012. Accuracy and precision of species trees: effects of locus, individual, and base pair sampling on inference of species trees in lizards of the *Liolaemus darwini* group (Squamata, Liolaemidae). *Systematic Biology*, 61: 272-288.
- Camargo A, Sites JJ. 2013. Species delimitation: a decade after the renaissance. *Agricultural and Biological Sciences "The Species Problem - Ongoing Issues"*, (ed.) Igor Ya. Pavlinov, Intech., ISBN 978-953-51-0957-0.
- Carstens BC, Knowles LL. 2007. Estimating Species Phylogeny from Gene-Tree Probabilities despite Incomplete Lineage Sorting: An Example from *Melanoplus* Grasshoppers. *Systematic Biology*, 56:400-411.
- Carstens BC, Pelletier TA, Reid NM, Satler JD. 2013. How to fail at species delimitation. *Molecular Ecology*, 22: 4369-4383.
- Carlström A. 1986. New taxa and notes from the SE Aegean area and SW Turkey. *Willdenowia*, 16:73 - 78.

- Chater AO, Walters SM, Akeroyd JR. 1993. *Silene L.* In: Tutin, T.G., Burges N.A., Chater A.O., Edmonson, J.R., Heywood, V.H., Moore, D.M., Valentine, D.H., Walters, S.M. & Webb, D.A. (eds.). *Flora Europaea 1*. Cambridge University Press., Cambridge, 1:191-218.
- Chowdhuri PK. 1957. Studies in the genus *Silene* Notes from the Royal Botanic Garden, Edinburgh, 22:221-278.
- Coode MJE, Cullen J. 1967. *Silene L.* In Davis, P.H.(ed.), *Flora of Turkey and the East Aegean Islands*. Edinburgh Univ. Press, Edinburgh, 2:179-242.
- Daly DC, Cameron KM, W. SD. 2001. Plant Systematics in the Age of Genomics. *Plant Physiology*, 127:1328-1333.
- Dávalos LM, Porzecanski AL. 2009. Accounting for molecular stochasticity in systematic revisions: species limits and phylogeny of *Paroaria*. *Molecular Phylogenetics and Evolution*, 53:234 - 248.
- Degnan JH, Rosenberg NA. 2006. Discordance of species trees with their most likely gene trees *PLoS Genetics*, 2:762-768.
- Degnan JH, Rosenberg NA. 2009. Gene tree discordance, phylogenetic inference and the multispecies coalescent. *Trends in Ecology & Evolution*, 24:332 - 340.
- De Queiroz K. 1998. The general lineage concept of species, species criteria, and the process of speciation: A conceptual unification and terminological recommendations. In *Endless Forms: Species and Speciation*, D. J. Howard and S. H. Berlocher (eds.). Oxford University Press, Oxford, England.
- De Queiroz K. 2007. Species concepts and species delimitation. *Systematic Biology*, 56:879-886.
- Desfeux C, Lejeune B. 1996. Systematics of Euromediterranean *Silene* (Caryophyllaceae): evidence from a phylogenetic analysis using ITS sequences. *Comptes Rendus de l'Academie des Sciences. Serie III, Sciences de la Vie (Paris)*, 319:351-358.
- Doyle JJ. 1992. Gene trees and species trees: molecular systematics as one-character taxonomy. *Systematic Botany*, 17:144-163.
- Drummond AJ, Ho SY, Phillips MJ, Rambaut A. 2006. Relaxed phylogenetics and dating with confidence. *PLoS Biology*, 4:e88.
- Drummond AJ, Rambaut A. 2007. BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evolutionary Biology*, 7:214.
- Drummond AJ, Suchard MA, Xie D, Rambaut A. 2012. Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Molecular Biology and Evolution*, 29:1969-1973.
- Edwards SV, Liu L, Pearl DK. 2007. High-resolution species trees without concatenation *PNAS*, 104:5936–5941.
- Edwards SV. 2009. Is a new and general theory of molecular systematics emerging? *Evolution*, 13: 1-19.
- Edwards DL, Knowles LL. 2014. Species detection and individual assignment in species delimitation: can integrative data increase efficacy? *Proceedings of the Royal Society B: Biological Sciences*, 281:20132765.
- Egge JJD, Simons AM. 2006. The challenge of truly cryptic diversity: diagnosis and description of a new madtom catfish (Ictaluridae: *Noturus*). *Zoologica Scripta*, 35:581-595.
- Eggen F. 2006. Systematics in *Sileneae* (Caryophyllaceae) - Taxonomy and Phylogenetic patterns. PhD. thesis.
- Ence DD, Carstens BC. 2011. SpedeSTEM: a rapid and accurate method for species delimitation. *Molecular Ecology Resources*, 11:473-480.
- Erixon P, Oxelman B. 2008. Reticulate or tree-like chloroplast DNA evolution in *Sileneae* (Caryophyllaceae)? *Molecular Phylogenetics and Evolution*, 48:313-325.

- Ewens WJ. 1979. *Mathematical Population Genetics*. Springer, New York.
- Ewing B, Hillier L, Wendl M, Green P. 1998. Base-calling of automated sequencer traces using phred. I. Accuracy assessment. *Genome Resources*, 8:175-185.
- Felsenstein J. 2006. Accuracy of coalescent likelihood estimates: do we need more sites, more sequences, or more loci? *Molecular Biology and Evolution*, 23:691-700.
- Felsenstein J. 2007. *Theoretical Evolutionary Genetics*. Genome 562 University of Washington.
- Frajman B, Eggens F, Oxelman B. 2009. Hybrid origins and homoploid reticulate evolution within *Heliosperma* (Sileneae, Caryophyllaceae) – a multigene phylogenetic approach with relative dating. *Systematic Biology*, 58:328-345.
- Fujita MK, Leaché AD. 2011. A coalescent perspective on delimiting and naming species: a reply to Bauer et al. *Proceedings of the Royal Society B: Biological Sciences*, 278:493-495.
- Fujisawa T, Barraclough TG. 2013. Delimiting Species Using Single-Locus Data and the Generalized Mixed Yule Coalescent Approach: A Revised Method and Evaluation on Simulated Data Sets. *Systematic Biology*, 62:707-724.
- Gadagkar SR, Rosenberg MS, Kumar S. 2005. Inferring species phylogenies from multiple genes: concatenated sequence tree versus consensus gene tree. *Journal of Experimental Zoology. Part B Molecular Developmental Evolution*, 304:64-74.
- Gelman A, Meng X. 1998. Simulating normalizing constants: from importance sampling to bridge sampling to path sampling. *Statistical Sciences*, 13:163 - 185.
- Grummer J, Bryson JR, Reeder T. 2014. Species Delimitation Using Bayes Factors: Simulations and Application to the *Sceloporus scalaris* Species Group (Squamata: Phrynosomatidae). *Systematic Biology*, 63:119-133.
- Greuter W. 1995. *Silene* (Caryophyllaceae) in Greece: a subgeneric and sectional classification. *Taxon*, 44:543 - 581.
- Greuter W, Burdet HM, Long G, et al. 2007. Med-Checklist: A critical inventory of vascular plants of the circum-mediterranean countries. <http://ww2.bgbm.org/mcl/home>
- Harrington RC, Near TJ. 2012. Phylogenetic and Coalescent Strategies of Species Delimitation in Snubnose Darters (Percidae: *Etheostoma*). *Systematic Biology*, 61:63 - 79.
- Hastings WK. 1970. Monte Carlo Sampling Methods Using Markov Chains and Their Applications. *Biometrika*, 57:97–109.
- Heled J, Bryant D, Drummond AJ. 2013. Simulating gene trees under the multispecies coalescent and time-dependent migration. *BMC Evolutionary Biology*, 13:44.
- Heled J, Drummond AJ. 2010. Bayesian inference of species trees from multilocus data. *Molecular Biology and Evolution*, 27:570-580.
- Hey J. 2004. What's So Hot about Recombination Hotspots? *PLoS Biology*, 2:e190.
- Hey J. 2006. Recent advances in assessing gene flow between diverging populations and species. *Current opinion in genetics & development*, 16:592-596.
- Hudson RR. 1990. Gene genealogies and the coalescent process. *Oxford Surveys Evolutionary Biology*, 7:1 - 144.
- Huelsenbeck J, Andolfatto P, Huelsenbeck E. 2011. Structurama: Bayesian inference of population structure. *Evolutionary Bioinformatics*, 7:55-59.
- Isaac NJB, Mallet J, Mace GM. 2004. Taxonomic inflation: its influence on macroecology and conservation. *Trends in Ecology & Evolution*, 19:464 - 469.
- Jennings WB, Edwards SV. 2005. Speciation history of Australian grass finches (*Poephila*) inferred from 30 gene trees. *Evolution*, 59:2033-2047.



- Jones G, Sagitov S, Oxelman B. 2013. Statistical inference of allopolyploid species networks in the presence of incomplete lineage sorting. *Systematic Biology*, 62:467.
- Jones G, Oxelman B. 2014. DISSECT: an assignment-free Bayesian discovery method for species delimitation under the multispecies coalescent. In review
- Kass RE, Raftery AE. 1995. Bayes factors. *Journal of American Statistic Association*, 90:773-795.
- Kingman JFC. 1982a. The coalescent. *Stochastic Processes and their Applications*, 13:235 - 248.
- Kingman JFC. 1982b. On the genealogy of large populations *Journal of Applied probability*, 19A: 27 - 43.
- Kosakovsky Pond SL, Posada D, Gravenor MB, Woelk CH, Frost SD. 2006. GARD: a genetic algorithm for recombination detection. *Bioinformatics*, 22:3096-3098.
- Kubatko LS, Carstens BC, Knowles LL. 2009. STEM: species tree estimation using maximum likelihood for gene trees under coalescence. *Bioinformatics*, 25:971-973.
- Kubatko L. 2009. Identifying Hybridization Events in the Presence of Coalescence via Model Selection *Systematic Biology*, 58:478-488.
- Kubatko LS, Gibbs HL, Bloomquist EW. 2011. Inferring species-level phylogenies and taxonomic distinctiveness using multilocus data in *Sistrurus rattlesnakes*. *Syst Biol*, 60:393 - 409.
- Lartillot N, Philippe H. 2006. Computing Bayes factors using thermodynamic integration. *Systematic Biology*, 55:195-207.
- Lazkov GA. 2003. The genus *Silene* (Caryophyllaceae) in the flora of Eurasia (systematics, distribution, history). Abstract of PH.D. thesis: 3-37.
- Lazkov GA. 2004. A new species of the genus *Silene* (Caryophyllaceae) from Iraq. *Botanicheskii Zhurnal* 89: 1180–1181.
- Leache AD, Fujita MK. 2010. Bayesian species delimitation in West African forest geckos (*Hemidactylus fasciatus*). *Proceedings of the Royal Society B: Biological Sciences*, 277: 3071- 3077.
- Leaché AD, Fujita MK, Minin VN, Bouckaert RR. 2014. Species Delimitation using Genome-Wide SNP Data. *Systematic Biology*, 63:534-542.
- Leaché AD, Rebecca BH, Rannala B, Yang Z. 2014. The Influence of Gene Flow on Species Tree Estimation: A Simulation Study. *Systematic Biology*, 63:17-30.
- Lewis H. 1966. Speciation in flowering plants. Rapid chromosome reorganization in Marginal populations is a frequent mode of speciation in plants. *Science*, 152:167-172.
- Lichten M, Goldman ASH. 1995. Meiotic recombination hotspots. *Annual Review of Genetics*, 29:423-444.
- Liu L, Pearl DK. 2007. Species trees from gene trees: reconstructing Bayesian posterior distributions of a species phylogeny using estimated gene tree distributions. *Systematic Biology*, 56:504-514.
- Liu L, Pearl DK, Brumfield RT, Edwards SV. 2008. Estimating species trees using multiple-allele DNA sequence data. *Evolution*, 62:2080-2091.
- Liu L. 2008. BEST: Bayesian estimation of species trees under the coalescent model. *Bioinformatics*, 24:2542 - 2543.
- Liu L, Yu L, Kubatko L, Pearl DK, Edwards SV. 2009a. Coalescent methods for estimating phylogenetic trees. *Molecular Phylogenetics and Evolution*, 53:320 - 332 328.
- Liu L, Yu L, Pearl DK, Edwards SV. 2009b. Estimating species phylogenies using coalescence times among sequences. *Systematic Biology*, 58:468 - 477.
- Maddison WP. 1997. Gene Trees in Species Trees. *Systematic Biology*, 46:523 - 536.
- Maddison WP, Knowles LL. 2006. Inferring phylogeny despite incomplete lineage sorting. *Syst Biol.*, 55:21-30.

- Marcussen T, Sandve SR, Heier L, Spannagl M, Pfeifer MP, IWGSCI., Jakobsen KS, Wulff BBH, Steuernagel B, Mayer KFX, *et al.* 2014. Ancient hybridizations among the ancestral genomes of bread wheat. *Science*, 345:no: 6194.
- Mayr E. 1942. *Systematics and the origin of species* New York, NY: Columbia University of Press.
- Martin DP, Lemey P, Lott M, Moulton V, Posada D & Lefevre P. 2010. RDP3: a flexible and fast computer program for analyzing recombination. *Bioinformatics*, 26: 2462-2463.
- Mayden RL. 1997. A hierarchy of species concept: the denouement in the saga of the species problem.
- Meikle RD. 1977. *Flora of Cyprus*, vol.1
- Melzheimer V. 1988. Caryophyllaceae In *Flora Iranica*. Rechinger KH ed. Akademische Druck-U. Verlagsanstalt, Graz, Austria, 163:353-508.
- Minin VN, Dorman KS, Fang F, Suchard MA. 2005. Dual multiple change-point model leads to more accurate recombination detection. *Bioinformatics*, 21:3034-3042.
- Miralles A, Vences M. 2013. New Metrics for Comparison of Taxonomies Reveal Striking Discrepancies among Species Delimitation Methods in Madascincus Lizards *PlosOne*, 8:e68242.
- Morton JK. 2005. *Silene*. In *Flora of North America* Editorial committee (eds.) *Flora of North America north of Mexico*. *Flora of North America north of Mexico*. Oxford University Press., New York, 5:166-214.
- Niemiller ML, Near TJ, Fitzpatrick BM. 2012. Delimiting species using multilocus data: diagnosing cryptic diversity in the southern cavefish, *typhlichthys subterraneus* (teleostei: amblyopsidae). *Evolution*, 66:846 - 866.
- Nordborg M. 2001. Coalescent Theory, In *Handbook of Statistical Genetics*, editing by Balding D, Bishop M, Cannings C, Wiley Chichester, UK, pp. 179-212.
- Olave M, Solà E, Knowles LL. 2014. Upstream analyses create problems with DNA-based species delimitation. *Systematic Biology*, 63:263-271.
- O'Meara BC. 2010. New heuristic methods for joint species delimitation and species tree inference. *Systematic Biology*, 59:59-73.
- Otth A. 1824. *Silene* L. de Candolle, A.-P. (ed.), *Prodromus systematis naturalis Regni Vegetabilis* de Candolle, A.-P. (ed.), *Prodromus systematis naturalis Regni Vegetabilis*, 1.
- Oxelman B, Lidén M. 1995. Generic boundaries in the tribe *Sileneae* (*Caryophyllaceae*) as inferred from nuclear rDNA sequences. *Taxon*, 44:525-542.
- Oxelman B, Lidén M, Berglund D. 1997. Chloroplast rps16 intron phylogeny of the tribe *Sileneae* (*Caryophyllaceae*) *Plant Systematics and Evolution*, 206:393- 410.
- Oxelman B, Greuter W. 1997. *Silene* section *Atocion* Otth, *Silene* section *Sedoideae* Oxelman and Greuter, *Silene* section *Rubellae* (Chowdhuri) Oxelman and Greuter ined., and section *Rigidulae* Schischkin. In Strid, A. and Tan, K.(eds.). *Flora Hellenica*, 1:248- 310.
- Oxelman B, Lidén M, Rabeler RK, Popp M. 2001. A revised generic classification of the tribe *Sileneae* (*Caryophyllaceae*). *Nordic journal of botany*, 20:515 - 518.
- Oxelman B, Rautenberg A, Thollesson M, Larsson A, Frajman B, Eggens F, Petri A, Aydin Z, Töpel M, Brandtberg-Falkman A. 2013. *Sileneae* taxonomy and systematics. Available from <http://www.sileneae.info>.
- Page RDM, Holmes EC. 1998. *Molecular Evolution: A Phylogenetic Approach*. Blackwell Science, Oxford, U.K.
- Petri A, Oxelman B. 2011. Phylogenetic relationships within *Silene* (*Caryophyllaceae*) section *Physolychnis*. *Taxon*, 60:953-968.

- Petri A, Pfeil BE, Oxelman B. 2013. Introgressive hybridization between Anciently diverged lineages of *Silene* (Caryophyllaceae). *PLoS One*, 8:e67729.
- Pimm SL, Russell GJ, Gittleman JL, Brooks TM. 1995. The future of biodiversity *Science*, 269:347-350.
- Pons J, Barraclough TG, Gomez-Zurita J, Cardoso A, Duran DP, Hazell S, Kamoun S, Sumlin WD, Vogler AP. 2006. Sequence-based species delimitation for the DNA taxonomy of undescribed insects. *Systematic Biology*, 55:595-609.
- Popp M, Oxelman B. 2004. Evolution of a rna polymerase gene family in *Silene* (Caryophyllaceae) incomplete concerted evolution and topological congruence among paralogues. *Systematic Biology*, 53:914-932.
- Rannala B, Yang Z. 2003. Bayes Estimation of Species Divergence Times and Ancestral Population Sizes Using DNA Sequences From Multiple Loci. *Genetics*, 164:1645-1656.
- Rannala B, Yang Z. 2013. Improved reversible jump algorithms for Bayesian species delimitation. *Genetics*, 194:245 - 253.
- Rautenberg A, Hathaway L, Oxelman B, Prentice HC. 2010. Geographic and phylogenetic patterns in *Silene* section *Melandrium* (Caryophyllaceae) as inferred from chloroplast and nuclear DNA sequences. *Molecular Phylogenetics and Evolution* 57: 978–991.
- Rasmussen MD, Manolis K. 2012. Unified modeling of gene duplication, loss, and coalescence using a locus tree. *Genome Research*, 22:755-765.
- Reid NM, Hird SM, Brown JM, Pelletier TA, McVay JD, Satler JD, Carstens BC. 2014. Poor Fit to the Multispecies Coalescent is Widely Detectable in Empirical Data. *Systematic Biology*, 63: 322-333.
- Rosenberg NA, Nordborg M. 2002. Genealogical trees, coalescent theory, and the analysis of genetic polymorphisms. *Nature Review of Genetics*, 3:380-390.
- Rosenberg NA, Tao R. 2008. Discordance of Species Trees with Their Most Likely Gene Trees: The Case of Five Taxa. *Systematic Biology*, 57:131-140.
- Satler JD, Carstens BC, Hedin M. 2013. Multilocus Species Delimitation in a Complex of Morphologically Conserved Trapdoor Spiders (Mygalomorphae, Antrodiaetidae, *Aliatypus*). *Systematic Biology*, 62 805 - 823.
- Scheen AC, Pfeil BE, Petri A, Heidari N, Nylinder S, Oxelman B. 2012. Use of allele-specific sequencing primers is an efficient alternative to PCR subcloning of low-copy nuclear genes. *Molecular Ecology Resources*, 12:128-135.
- Sibley CG, Ahlquist JE. 1990. *Phylogeny and classification of birds: a study in molecular evolution* New Haven: Yale University Press.
- Sites JW, Marshall JC. 2004. Operational Criteria for Delimiting Species. *Annual Review of Ecology, Evolution, and Systematics*, 35:199-227.
- Soltis DE, Soltis PS. 2003. The Role of Phylogenetics in Comparative Genetics1. *Plant Physiology* August 2003 132:1790-1800.
- Swofford DL. 2003. PAUP\*: phylogenetic analysis using parsimony, version 4.0b10.
- Staden R. 1996. The staden sequence analysis package. *molecular Biotechnology*, 5:233-241.
- Takahata N. 1989. Gene genealogy in three related populations: consistency probability between gene and population trees. *Genetics*, 122:957 – 966.
- Takahata N. 1991. Genealogy of Neutral Genes and Spreading of Selected Mutations in a Geographically Structured Population. *Genetics*, 129:585 - 595.
- Thomas CD, Cameron A, Green RE, Bakkenes M, Beaumont LJ, Collingham YC, Erasmus BFN, Ferreira de S, M., Grainger A, Hannah L, *et al.* 2004. Extinction risk from climate change. *Nature*, 427:145-148.

- Wiens JJ. 1998. Combining Data Sets with Different Phylogenetic Histories. *Systematic Biology*, 47:568-581.
- Wiens JJ, Servedio MR. 2000. Species delimitation in systematics: inferring diagnostic differences between species. *Proceedings of the Royal Society B: Biological Sciences*, 267:631-636.
- Wiens JJ. 2007. Species delimitation: new approaches for discovering diversity. *Systematic Biology*, 56:875 - 878.
- Xie W, Lewis P, Fan Y, Kuo L, Chen M-H. 2011. Improving marginal likelihood estimation for Bayesian phylogenetic model selection. *Systematic Biology*, 60:150-160.
- Yang Z, Rannala B. 2010. Bayesian species delimitation using multilocus sequence data. *Proceedings of the National Academy of Sciences of the United States of America*, 107:9264-9269.
- Yildiz K, Gücel S. 2006. Chromosome numbers of 16 endemic plant taxa from northern Cyprus. *Turkish Journal of Botany*, 30:181-192.
- Yildiz K, Gücel S, Dadandi MY. 2009. A palynological investigation of endemic taxa from northern Cyprus. *Pakistan Journal of Botany*, 41:991-1007.
- Zhang C, Zhang D, Zhu T, Yang Z. 2011. Evaluation of a Bayesian Coalescent Method of Species Delimitation. *Systematic Biology*, 60:747-761.