Large scale plastomics approaches for the study of evolution and adaptive signatures in angiosperms

Dissertation der Fakultät für Biologie
der Ludwig-Maximilians-Universität München

vorgelegt von
Huan Li

München, 2017

Diese Dissertation wurde angefertigt
unter der Leitung von Prof. Dr. Dario Leister
im Bereich von Fakultät für Biologie
an der Ludwig-Maximilians-Universität München

Erstgutachter: Prof. Dr. Dario Leister
Zweitgutachter: Prof. Dr. Wolfgang Frank

Tag der Abgabe: 24 August 2017
Tag der mündlichen Prüfung: 9 November 2017

Erklärung

Ich versichere hiermit an Eides statt, dass meine Dissertation selbständig und ohne unerlaubte Hilfsmittel angefertigt worden ist.
Die vorliegende Dissertation wurde weder ganz, noch teilweise bei einer anderen Prüfungskommission vorgelegt.
Ich habe noch zu keinem früheren Zeitpunkt versucht, eine Dissertation einzureichen oder an einer Doktorprüfung teilzunehmen.

München, den 24 August 2017 Huan Li

Summary

Chloroplasts are the centers of photosynthesis in plant cells and harbor their own genome. Until recently, the study of the organization and evolution of chloroplast genomes was limited by the dearth of sequence information. Nowadays, the next generation sequencing (NGS) technology provides the possibility to make use of large sets of sequencing data in phylogenetic and evolutionary studies, especially for chloroplast DNA. However, the molecular mechanisms of plastid genome evolution and its relevance for environmental adaptation remain poorly understood. This study contributes to filling this gap by providing and analyzing a large number of protein coding sequences (CDS) of chloroplasts in flowering plants. More than 1,100 representatives of wild species from an alpine flora were subjected to sequencing of high-copy number genes using a shotgun approach on genomic DNA. The major results obtained are:

1) At the genus level, the phylogenetic reconstruction of 34 taxa of *Aquilegia* provided a basic assessment of the relationships among Eurasian taxa through two datasets. Thereby, chloroplast genes with lower guanine-cytosine (GC) content and GC content at the third position of codon encoded genes with higher amino acid polymorphisms.

2) At the family level, 71 chloroplast CDS from 95 species in the Brassicaceae family were used to reconstruct their phylogenetic relationships and analyze the patterns of molecular evolution characterizing them. A total of 33 genes were found to be under positive selection. Overall, three major phylogenetic lineages (I-III) were defined in this study.

3) Different computational approaches were used for detection of natural selection in 11 families. The chloroplast CDSs, analyzed family by family, showed patterns of molecular evolution consistent with increased rates of average positive selection for a limited number of genes. Among them, *rbcL* was inferred as being under positive selection in 8 of the 11 families considered, indicating recurrent functional selection. A total of 46 branches in the 11 family-level phylogenies showed patterns of molecular evolution consistent with positive selection, and in general the patterns of selection were mostly consistent with directional rather than relaxed selection. Five selected families were analyzed to test whether the ratio between non-synonymous and synonymous nucleotide changes significantly differed between low and high altitude taxa, and four genes showed patterns of positive selection differentially associated with altitude. Besides, this study further assessed whether these results might be affected by RNA-editing.

Taken together, this thesis provides a successful large-scale application of NGS technology for the elucidation of molecular evolution patterns at different levels of angiosperms phylogeny, and offers a new view on the structural and functional features of the chloroplast genome and its relationships with adaptation in wild plant species.

Zusammenfassung

Chloroplasten sind die Zentren der Photosynthese in Pflanzenzellen und besitzen ein eigenes Genom. Bis vor kurzen wurde die Untersuchung der Organisation und Evolution von Plastiden-Genomen durch die eingeschränkte Verfügbarkeit genetischer Information limitiert. Heutzutage bietet die "Next Generation Sequencing (NGS)" Technologie die Möglichkeit, große Mengen von Sequenzierungsdaten für phylogenetische und evolutionäre Studien zu nutzen, insbesondere von Chloroplasten-DNA. Allerdings sind die molekularen Mechanismen, die die plastidäre Genomevolution vorantreiben, und deren Relevanz für Adaptionsvorgänge weitgehend unverstanden. Das Ziel dieser Arbeit war diese Wissenslücke durch die Identifizierung und Analyse der Protein-kodierenden Sequenzen (CDS) von Chloroplasten aus Blütenpflanzen zu schließen. Mit Hilfe einer "Shot Gun" Genomsequenzierung wurden von mehr als 1.100 repräsentativen Wildarten aus einer alpinen Flora die Gene mit hoher Kopienzahl sequenziert. Die wichtigsten Ergebnisse waren:

1) Durch die phylogenetische Rekonstruktion von 34 Taxa von Aquilegia mit Hilfe von zwei Datensätzen konnte auf der Gattungsebene die Beziehungen zwischen eurasischen Taxa grundlegend bewertet werden. Dabei kodieren chloroplastidäre Gene mit geringem Guanin-Cytosin (GC)-Gehalt und geringerem GC-Gehalt an der dritten Position des Codons für Proteine mit erhöhtem Anteil von Aminosäurepolymorphismen.

2) Auf Familienebene wurden 71 Chloroplasten CDS von 95 Arten in der Familie Brassicaceae verwendet, um ihre phylogenetischen Beziehungen zu rekonstruieren und die Muster der zugrundeliegenden molekularen Evolution zu analysieren. Dabei wurden insgesamt 33 Gene gefunden, die positiver Selektion ausgesetzt sind. Insgesamt wurden in dieser Studie drei große phylogenetische Linien (I-III) definiert.

3) Verschiedene bioinformatische Ansätze wurden verwendet, um natürliche Selektion bei 11 verschiedenen Familien nachzuweisen. Jede Familie wurde bezüglich ihrer chloroplastidären CDS untersucht und eine Reihe von Gene mit erhöhter positiver Selektion gefunden. Darunter war das *rbcL* Gen, das bei 8 der 11 Familien positiver Selektion ausgesetzt ist, was auf eine wiederkehrende funktionale Selektion hinweist. Insgesamt 46 Äste in den 11 Familien-Phylogenien zeigten Muster molekularer Evolution auf, die mit positiver Selektion übereinstimmten, und im Allgemeinen waren die Selektionsmuster meist im Einklang mit gerichteter Selektion. Fünf ausgewählte Familien wurden daraufhin untersucht, ob das Verhältnis zwischen nicht-synonymer und synonymer Nukeotidaustausche sich signifikant zwischen Taxa aus unterschiedlicher Höhenlagen unterscheidet und es wurden vier solcher Gene gefunden. Außerdem wurde untersucht, ob diese Ergebnisse durch RNA-Editierung beeinflusst werden könnten.

Zusammengefasst leistet diese Arbeit eine erfolgreiche groß angelegte Anwendung der NGS-Technologie zur Aufklärung von molekularen Evolutionsmustern auf

verschiedenen Ebenen der Angiospermen-Phylogenie und bietet neue Einsichten in strukturelle und funktionellen Merkmale des Chloroplastengenoms und seiner Beziehungen zur Adaption bei Wildpflanzen.

Contents

List of figures

List of tables

Abbreviations

AIC	Akaike Information Criterion
aLRT	approximate likelihood ratio test
APG	Angiosperm Phylogeny Group
ATP	adenosine triphosphate
BI	bayesian inference
CDS	coding sequences
dN	non-synonymous substitutions
DOGMA	Dual Organellar GenoMe Annotator
dS	synonymous substitutions
FDR	false discovery rate
GC	guanine-cytosine
GTR	generalised time-reversible
IR	inverted repeat
IQR	interquartile range
LSC	large single copy
LSU	large subunit
ML	maximum likelihood
MP	maximum parsimony
mya	million years ago
NADPH	nicotinamide adenine dinucleotide phosphate
NCBI	National Center for Biotechnology Information
NGS	next generation sequencing
PSI	photosystem I
rRNA	ribosomal RNA
SSC	small single copy
SSU	small subunit
tRNA	transfer RNA

1. Introduction

1.1 The angiosperms

1.1.1 Description of angiosperms

The angiosperms (or flowering plants) are also known as Magnoliophyta. They are the most diverse group of land plants, with 416 families, approximately 13,164 known genera and a total of 295,383 known species (Christenhusz and Byng, 2016). Etymologically, angiosperm means a plant which can produce seeds within an enclosure, in other words, a fruiting plant. Angiosperms diverged from gymnosperms in the Triassic Period, in the time range between 245 to 202 million years ago (mya), and the first fossil record of angiosperms, *Archaeanthus linnenbergeri,* is known from 160 mya (Dilcher and Crane, 1984). They diversified extensively during the Lower Cretaceous, became widespread by 120 mya, and replaced conifers as the dominant trees from 100 to 60 mya.

Angiosperms have some distinctive features that distinguish them from gymnosperms including flowering organs, stamens with two pairs of pollen sacs, reduced male/female gametophyte, closed carpel enclosing the ovules, endosperm, etc. For instance, the flowers serve as the reproductive organs for the plant, providing them a means of sexually exchanging genetic information. Angiosperms have small pollen grains that spread genetic information from flower to flower, and these grains are much smaller than the gametophytes, or reproductive cells, used by non-flowering plants. Besides, a great advantage for angiosperms is the production of endosperm, which is a material that forms after fertilization and serves as a highly nutritional food source for the developing embryo and seedling.

The APG system (Angiosperm Phylogeny Group system) of plant classification is the first version of a modern, mostly molecular-based, system of plant taxonomy. The number of families in APG (1998) was 462, in APG II (2003) at maximum it was 457, in APG III (2009) it was 413, and APG IV (2016) it was adjusted to 416. There are eight main groups of living angiosperms. The basal angiosperms contain *Amborella*, Nymphaeales, Austrobaileyales, while core angiosperms include Chloranthales, Magnoliids, *Ceratophyllum*, eudicots and monocots (Palmer et al., 2004). The basic angiosperm form is woody or herbaceous, and woody forms (generally trees and shrubs) are rich in secondary tissues while herbaceous forms (herbs) rarely have any. The core angiosperms contain about 99.95% of the angiosperms species. Actually, eudicots and monocots are the largest and most diversified, with about 75% (210,008 known species) and 23% (74,273 known species) of angiosperm species, respectively.

1.1.2 Important genera and families in angiosperms for this study

1.1.2.1 The *Aquilegia* genus

The lower eudicot genus, *Aquilegia* (or columbine), is a perennial plant belonging to the family Ranunculaceae. It has approximately 70 species, which are widely distributed at higher latitudes throughout the temperate region of the northern hemisphere (North America, Europe and Asia) over the past 1 to 5 million years (Tamura, 1993; Kim et al., 2009; Kramer, 2009; Ballerini and Kramer, 2011). According to Hodges' researches, *Aquilegia* radiated initially in Eurasia and then expanded to North America (Hodges, 1997; Hodges et al., 2004). The *Aquilegia* genus shares a fairly simple, herbaceous body plant. The seedlings of *Aquilegia* have two cotyledons followed by helical and compound leaves. The mature petioles are usually quite long and leaflets are arranged in bifid or trifid units. The flowers of *Aquilegia* have five organ types arranged in pentamerous whorls and four to six free carpels in the center of the flower (Kramer, 2009). Additionally, *Aquilegia* has a seven-chromosome genome (S.A. Hodges, personal communication) which is about 300Mbp (Kramer, 2009; Ballerini and Kramer, 2011). Previous studies have demonstrated that the species within the *Aquilegia* genus show significant differences in vegetative and floral traits with specializations for different ecological niches and pollinators (Kramer, 2009; Kramer and Hodges, 2010). For example, the diversification of *Aquilegia* in North America is associated mainly with the adaptation to a number of different pollinators (Hodges, 1997; Fulton and Hodges, 1999). Additionally, spur length in the *Aquilegia* genus was found to have progressively increased, adaptively evolving to fit pollinator shifts in North American *Aquilegia* taxa (Whittall and Hodges, 2007).

1.1.2.2 The Brassicaceae family

Brassicaceae, or *Cruciferae*, is a medium-sized and economically relevant family with approximately 372 genera and 4,060 species distributed worldwide (Bremer et al., 2009). The largest genera are *Draba* (440 species), *Erysimum* (261 species), *Lepidium* (234 species), *Cardamine* (233 species), and *Alyssum* (207 species). This family belongs to the order Brassicales, and consists mostly of herbaceous plants with annual, biennial, or perennial lifespans. All continents except Antarctica are potential habitats, and most of the species are found concentrated in temperate regions of the northern hemisphere. For instance, the Mediterranean, Iran-Turanian, as well as Northwest America are regions where Brassicaceae are most widely distributed. This distribution has revealed a potential Iran-Turanian origin of Brassicaceae, a place where the family possibly originated and then spread to the other parts of the globe (Franzke et al., 2009). Besides, some genera

are commonly found in the southern hemisphere, such as *Draba* and *Lepidium*. Tropics and subtropical regions, mountainous, and alpine regions are also habitats where Brassicaceae taxa can often be found. Additionally, Brassicaceae contains many plants of economic importance that have been extensively altered and domesticated by humans, such as cabbage, broccoli, kale, kohlrabi, napa cabbage, turnip, and rutabaga. Other important agricultural crops in the family include horseradish, radish, and white mustard. The most important and universally studied species, *Arabidopsis thaliana*, is the model system species used for plant study in the world.

1.1.2.1 Other important families

Eight eudicots (Apiaceae, Asteraceae, Caryophyllaceae, Fabaceae, Lamiaceae, Plantaginaceae, Ranunculaceae, Rosaceae) and three monocots (Cyperaceae, Orchidaceae and Poaceae) clades were selected as families to study in depth the patterns of chloroplast evolution in this study. Apiaceae is a family of mostly aromatic flowering plants named after the type genus *Apium* and commonly known as the celery, carrot or parsley family. Asteraceae is a very large and widespread family of flowering plants with 1,911 genera and 13 subfamilies. It is important in herbal medicine, including *Grindelia*, yarrow, and many others. Caryophyllaceae is a large family with 81 genera and about 2,625 known species, which are distributed in temperate climates, with a few species growing on tropical mountains. Fabaceae is widely distributed, and is the third-largest land plant family with about 751 genera and 19,000 known species (Christenhusz and Byng, 2016). Lamiaceae is a family commonly known as the mint or deadnettle family with about 236 genera and 7,200 species. Plantaginaceae is a family of flowering plants in the order Lamiales (like Lamiaceae) with 4 genera and about 1,900 species. Ranunculaceae is a family with over 2,000 known species in 43 genera distributed worldwide, which contains the largest genera, *Ranunculus* (600 species), *Delphinium* (365), *Thalictrum* (330), *Clematis* (325), and *Aconitum* (300). Rosaceae, the rose family, is a medium-sized family including 4,828 known species in 91 genera, such as apples, pears, quinces, apricots, plums, cherries, peaches, etc. Cyperaceae is a monocotyledonous graminoid family known as sedges, containing around 90 genera and 5,500 species. Poaceae is another monocotyledonous family in the Poales order (like Cyperaceae) known as grasses. Orchidaceae is a diverse and widespread family with 763 genera and about 28,000 species. The family encompasses about 6–11% of all seed plants and the largest genera are *Bulbophyllum* (2,000 species), *Epidendrum* (1,500 species) and *Dendrobium* (1,400 species).

1.2 The chloroplast

1.2.1 Description of the chloroplast

Chloroplasts are organelles in plant and algal cells, which capture the energy from sunlight and convert it to adenosine triphosphate (ATP) and nicotinamide adenine dinucleotide (NADPH) molecules through the light reactions of photosynthesis. Although photosynthesis is regularly recognized as their key function, chloroplasts are capable of performing other specialized functions, including amino acid synthesis, fatty acid synthesis, nitrogen and sulfur assimilation, and interaction with the plant's environment (responses to heat, drought, salt, light, and so on). The chloroplast is a type of plastid, which is characterized by its high concentration of chlorophyll. Other types of plastids, such as chromoplast and leucoplast, contain little chlorophyll and do not carry out photosynthesis. Mohl first described the chloroplast in 1835 as discrete bodies within the green plant cell (Mohl, 1835). In 1884, Strasburger adopted the term "chloroplasts" (Strasburger, 1884), which is currently still in use.

It is generally accepted that chloroplasts originated from cyanobacteria through endosymbiosis approximately 1.5 billion years ago (Chan et al., 2011). In 1970, Margulis proposed the endosymbiotic hypothesis of the origin of eukaryotic cells (Margulis, 1970). According to this hypothesis, some primitive eukaryotic cells with phagocytic capacity engulfed some prokaryotic cyanobacteria. Since the bacteria had not been decomposed and digested, they gradually passed from parasitism to symbiosis and became organelles of the host cells. However, the origin of chloroplasts has raised many controversies. First, some researchers believe that it occurred as a single event in plant evolution (Cavalier-Smith, 2000), while others think it evolved multiple times (Whatley and Whatley, 1981). In other words, the phagocyted prokaryotes could be other algae or bacteria besides cyanobacteria. For instance, Reith and Munholland have studied the chloroplast genes of *Porphyra purpurea* (Reith and Munholland, 1995). Although they believe that the chloroplast of the algae is monophyletic, they found that it contains more than 70 genes which are not in land plants and green algae. This clearly explains why the origin of chloroplast is still doubtful. Evidences show that chloroplast DNA between land plants and green algae is relatively conserved (Ohta et al., 2003; Glöckner et al., 2000; Hallick et al., 1993), and this demonstrates that land plants evolved from green algae. Second, some researchers believe that chloroplasts in some plants were recruited more than once through successive endosymbiotic associations. The conjecture of secondary endosymbiosis begun with the discovery of an envelope wrapped outside a traditional chloroplast membrane. Gilson and his colleagues reported two independent endosymbionts in two model species, *B. natans* and *Cryptomonad sp.*, and speculated that chloroplasts of these two species were from green algae and red algae, respectively

(Gilson et al., 2006). Furthermore, Rogers et al. sequenced the chloroplast genome of *B. natans* and compared it with the one of green algae, demonstrating that the chloroplast of this species was obtained independently from those of other plants (Rogers et al., 2007). Therefore, the author inferred that this process took place at least twice through engulfment of endosymbionts.

Although chloroplasts are often found in collenchyma tissue, the plant cells containing chloroplasts are usually parenchyma cells (Roberts, 2007). A typical chlorenchyma cell of land plant contains about 10 to 100 chloroplasts. In most plants, chloroplasts are mostly concentrated in the leaves, but they are found in highest amounts in the stems in few plants, such as cacti. In addition, chloroplasts are highly dynamic because they move around within the plant cell, and occasionally split in two to reproduce. Actually, chloroplasts in plant and algal cells orient themselves to the best direction suitable for the harvesting of the available light. In higher plants, chloroplast movement is run by phototropins, which mediate avoidance from excessive blue light and enable the maximal exposure to red light to maximize absorption of light (Takagi, 2003). Most chloroplasts do not develop directly from proplastids or etioplasts in a photosynthetic cell. The proplastids differentiate into chloroplasts, and divide to create almost 30–70 chloroplasts. And chloroplasts are usually inherited from a single parent, only few flowering plants have biparental inheritance (Hansen et al., 2007). Actually, chloroplasts in most of angiosperm derive exclusively from maternal inheritance, but are transmitted by paternal inheritance in gymnosperms, such as *Larix* (Szmidt et al., 1987).

1.2.2 General characteristics of the chloroplast genome

The chloroplast has its own genome, and its DNA is a double-stranded circular molecule. Only in few cases the plastomes are linear molecules, like in *Acetabularia* where the isolated chloroplast DNA appears in “linear” and “looped” forms (Werz and Kellner, 1968). The chloroplast genome accounts for approximately 10% ~ 20% of the total plant DNA and has a size of around 120 ~ 210 kb (120-170kb in land plants) (Olejniczak et al., 2016). At present, the smallest genomes known are those of *Helicosporidium* sp. *Simulium jonesie* (37.4 kb) and *Ostreococcus tauri* (71.6 kb), while the largest ones are *Nephroselmis olivacea* (200.8 kb) and *Dunaliella salina* (269 kb) (Smith et al., 2010). By gene mapping and gene sequencing, it is confirmed that chloroplast genomes are generally conserved in terms of gene content and overall structure, and most of the genomes characterized so far have the highly conserved tetrad structure (Jansen et al., 2005): a large single copy (LSC, about 81-87kb) and a small single copy (SSC, 18-20kb) separated by a pair of inverted repeats (IRa and IRb, are about 22-28 kb). The two inverted repeat regions contain the same genes, but they are arranged in the opposite direction. Only few plants have a special genome structure due to the complete loss of

one inverted repeat sequence, such as *Trifolium subterraneum*, *Medicago truncatula* and *Cicer arietinum* (Cai et al., 2008). The inverted repeat sequence is the most conserved region encoding ribosomal RNA (rRNA) and some transfer RNA (tRNA), while the large unique sequence is the most variable.

All chloroplast genomes include protein-encoding genes, as well as tRNA and rRNA genes. The protein-encoding genes are mainly divided into three categories according to their functions: the first category of genes are related to transcription and translation; the second category is associated with photosynthesis; the third group includes factors for biosynthesis of amino acids, fatty acids, pigments, etc. It is particularly interesting that some hypothetical conserved reading frames that are associated with photosynthesis (ycf) were discovered only relatively recently. The *ycf3* and *ycf4* genes are encoding proteins necessary for the assembly or stability of photosystem II; the *ycf9* gene also takes part in photosynthetic function, besides, as it forms a unit with PsbC and PsbD proteins; the *ycf8* gene encodes the PsbT protein, a subunit of photosystem II in *Chlamydomonas reinhardtii* (Rochaix, 2005). The *ycf1* gene encodes Tic24 which is necessary for importing proteins into chloroplast in *Arabidopsis thaliana* (Kikuchi et al., 2013). Currently, the function of the *ycf2* gene remains unknown.

Recent studies demonstrated that genome size and GC content differentiation are the most important factors in chloroplast evolutionary processes in plants (Caetano-Anollés, 2005; Johnston et al., 2005; Barow and Meister, 2002; Šmarda et al., 2008). The variations of chloroplast genome size are mainly caused by different lengths of intergenic regions, analogously to what observed in the case of the mitochondrial genome. Higher plant chloroplast genomes usually contain 14-18 split genes (i.e., genes containing introns), and the total number of introns is almost 20. But most chloroplast genes possess only one intron which is generally conserved in structure, and long noncoding regions and two introns are rare. Thus, introns have much smaller effect on genome size in chloroplast genes than in mitochondrial genes (Odintsova and Yurina, 2005). Although introns are conserved in the chloroplast genome, intron losses have been reported in several species: *Bambusa* sp. (Wu et al., 2009) and *Cicer arietinum* (Jansen et al., 2008). These genes are an ATP synthase (*atpF*), a Clp protease (*clpP*), an RNA polymerase (*rpoC2*), and ribosomal protein (*rpl2, rps12* and *rps16*) subunits (Jansen et al., 2007). The content of GC pairs is a trait strongly distinguishing chloroplast genomes from nuclear genomes. The GC content of plastomes is about 30%-40%, and this peculiar composition is the result of multiple chloroplast traits, like the bacterial-like properties of their DNA polymerase and DNA repair systems (Nielsen et al., 2010). The GC content of mitochondrial genome is higher than that of the chloroplast genomes in land plant, such as melon (Rodríguez-Moreno et al., 2011), but the functional relevance of this finding is still not clear.

There are many significant advantages for using chloroplast genomes in phylogeny studies, such as their sufficiently small genome size, the conserved single copy region and the moderate rate of nucleic acid replacement relative to nuclear DNA. In addition, coding regions and non-coding regions in chloroplast have obviously different rates of molecular evolution, so plastome sequences can be applied to elucidate evolutionary questions at different levels of systematic research.

1.2.3 Advance in chloroplast genome sequencing technology

The development of whole-genome sequencing technology for individual organism constitutes a landmark in biology, because since then sequence information is widely used in molecular evolution and phylogeny studies. Actually, improvement in sequencing technology is one of the most important factors which caused the rapid advancement in chloroplast genomics. The first generation sequencing is Sanger sequencing, which was developed by Frederick Sanger in 1977. The second generation sequencing technology is a general term for sequencing techniques that are modified on the basis of the synthesis termination method, ligation sequencing, pyrosequencing and hybridization sequencing. The traditional, second-generation sequencing platforms include 454 sequencing platform (Roche), Solexa sequencing (Illumina), and SOLiD sequencing platform (ABI, Applied Biosystems) (Shendure and Ji, 2008). Moore and his colleagues firstly sequenced chloroplast genome of *Nandina domestica* and *Platanus occidentalis* by using the 454 sequencing platform (Moore et al., 2006). Cronn et al. simultaneously sequenced eight chloroplast genomes of conifers through the Solexa sequencing platform (Cronn et al., 2008), and proved that the Solexa sequencing platform could obtain chloroplast genome data efficiently and rapidly. With the advent of second generation sequencing technology, genomics and functional genomics entered into a low cost era, characterized by large scale and high-throughput sequencing. This, in turn, brought novel opportunities for chloroplast genome whole sequencing and phylogeny. Due to the moderate size, conserved gene content and genome structure of chloroplast genomes, and the fact that single molecules of the chloroplast genome rarely display variation, the second generation sequencing technology is extremely suitable for the chloroplast genome. Illumina is the major platform currently utilized for chloroplast genome as it allows the use of rolling circle amplification products (Daniell et al., 2016). Compared to Sanger sequencing, second generation sequencing platforms support much higher amounts of sequencing data. However, a third-generation sequencer, PacBio system, now uses single molecule real time (SMRT) sequencing for chloroplast genomes (Ferrarini et al., 2013; Stadermann et al., 2015). Its reads length is longer than second-generation sequencing platforms, but it still has low accuracy in raw data. While the market is still dominated by Illumina for plastome sequencing, the combined use of second and third generation

sequencing provides the advantages of high throughput, low cost and high accuracy of second generation technologies with the increased length and speed of third generation methods. The first examples of plastomes sequenced by a combination of Illumina and PacBio sequencing are appearing (e.g. *Potentilla micrantha* chloroplast genome; Ferrarini et al., 2013), and possibly this approach will be increasingly used in future.

With the development of large-scale sequencing technology, the study of chloroplast genomes has been dramatically simplified and deepened. The chloroplast genome research not only potentially helps to improve chloroplast function to breed new valuable varieties through genetic transformation systems, but also reveals the mechanisms of photosynthesis at the molecular level, nuclear-cytoplasm interactions, and origin and evolution of species (Maliga, 1993). Since the first chloroplast genome of tobacco, *Nicotiana tabacum*, was sequenced in 1986 (Shinozaki et al., 1986), in less than 30 years the number of sequenced chloroplast genomes grew exponentially. In recent years, public sequence databases (such as NCBI, Genbank) currently record over 1000 species of whole chloroplast genome and some partial chloroplast genome sequences (Figure 1), but most of them are from land plants, Viridiplantae. Additionally, there are some species which contain different subspecies or different cultivars for which several chloroplast genomes have been obtained, such as rice, including *O. sativa* subsp. *indica* (Tang et al., 2004) and *O. sativa* subsp. *japonica* (Hiratsuka et al., 1989). The rapid progress made by chloroplast genome research in the last decade is destined to increase even further in the years to come, so one can state without doubts that the golden era of plastome genomics has already started (Shaw et al., 2014).

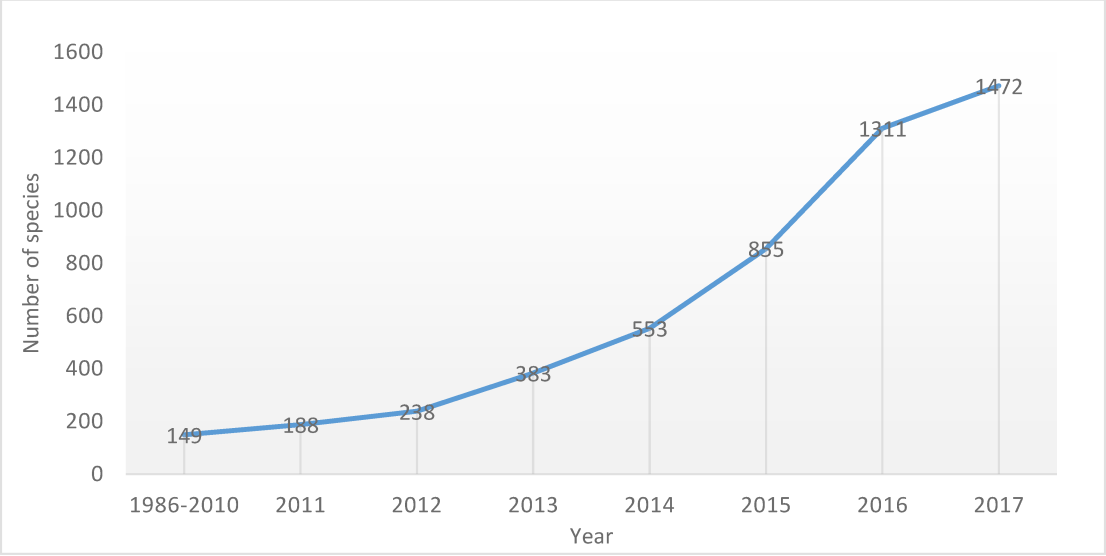

Figure 1.1: Number of chloroplast genome records in NCBI. X-axis presents every year the number of whole chloroplast genome records in NCBI database listed in Y-axis.

1.3 Phylogenetic studies of plastomes

1.3.1 Systematic phylogeny studies in angiosperms

The forms of life on Earth are diverse, but they all have a common evolutionary history. Distinct understanding of phylogenetic relationships among different biological groups is not only the premise of evolutionary biology research and the foundation of taxonomy and denominating, but also the basis of many other branches of biological research. So, constructing a reliable phylogenetic tree is not only the focus of phylogenetic research, but also one of the most important parts of biological research in general. However, reconstructing the phylogenetic relationships in angiosperms is extremely challenging because of the ancient age of the clades, the extinction of major lineages, and the extreme molecular rates (Kenrick and Crane, 1997; Doyle, 2006; Rothfels et al., 2012; Soltis et al., 2002; Smith and Donoghue, 2008). Traditionally, the most extensive analyses of angiosperms relationships based on nuclear gene sequence data have relied largely on 18S/26S rDNA sequences (Chaw et al., 1997; Soltis et al., 1999; Nickrent et al., 2000). Some studies also utilized mitochondrial gene sequence data, either alone or in combination with other data (Duff and Nickrent, 1999; Qiu et al., 2006). Many studies have recently proven that chloroplast genomic data are very useful for helping resolve plant phylogeny and evolution (Jansen et al., 2007; Moore et al., 2007; Zhong et al., 2014).

Several techniques have been employed for plant phylogeny using chloroplast DNA, from the less to the most informative: 1) DNA hybridization technology. This technology provides information on the whole genome, but it is work-intensive and difficult to standardize. 2) Single-strand conformation polymorphism (SSCP) analysis. It is fast, but not very informative. It provides information on the portions of DNA amplified by PCR. 3) Restriction fragment length polymorphism (RFLP) analysis. It is derived from the variation of genomic DNA which has the characteristics of stable inheritance and specificity, and obtains a large number of polymorphisms reflecting the genetic difference of species to study phylogenetic relationships, phylogeny and evolution of plant groups, especially between genera, species and even varieties. 4) Microsatellite sequence analysis. It is commonly used in distinguishing closely related genotypes. 5) Nucleotide sequence analysis. This technique is the most comprehensive and thorough comparison of biological genetic material and can be used for the study of genetic relationship at any level (including family, genus, species, etc.).

With twenty years of development, molecular systematics research has started from single genes, passing from combinations of multiple genes to reach the use of entire organelle genomes. In the angiosperm chloroplast*, rbcL, atpB, ndhF* and *matK* are the most common genes used as marker genes for traditional phylogenetic studies (Kim and

Jansen, 1995; Soltis et al., 2000; Hilu et al., 2003). The length of *atpB* and *rbcL* genes is about 1,500 bp, and the nucleotide substitution rate is around 0.068-0.108 (substitution site/unit time). The *ndhF* gene (around 2000 bp) is longer than *rbcL* and its evolutionary rate is faster. In addition, the evolutionary rate of *matK* gene (≥1,500 bp) is the fastest among chloroplast genes, and this gene is more suitable for studying the phylogenetic relationships among closely related species. However, phylogenetic trees based on different single genes are often different in angiosperms, and the support values of branches are not high. This is mainly due to the short sequences of single genes and the relatively few informative sites they contain (Delsuc et al., 2005; Jeffroy et al., 2006). Although through combination of several monogenic sequences the phylogenetic relationships among angiosperms have gained clearer and higher support, more and more chloroplast genomes have been sequenced and their use has gradually become the trend in the phylogeny of angiosperms (Moore et al., 2010). At present, the modern phylogenetic framework of angiosperms has been constructed by collaborative efforts of many molecular taxonomists around the world. For example, the APG classification system has been recently improved and updated to APG IV (2016). Besides, non-coding regions (introns and intergenic regions) of chloroplast have been used more and more often in phylogenetic analysis of plants (Small et al., 1998). Compared with coding genes, non-coding regions are less restricted in their function, and they have faster evolutionary rates either in nucleic acid substitutions or the accumulation of insertion/deletion mutations. Moreover, non-coding regions provide more systematic information than coding genes, so they are usually applied to the study of lower level taxa.

The variation of coding regions in chloroplast is normally associated to large phenotypic variations, and their slower evolutionary rate is useful for phylogenetic study of higher taxonomic groups, like families and orders. On the other hand, the mutation of non-coding regions has usually little effects on plants phenotype, and it is commonly used for lower taxonomic level phylogenetic reconstructions, like at the species and genus level, due to faster evolutionary rate. However, other DNA sequences (mitochondrial DNA and nuclear DNA) have been applied widely to phylogenetic reconstruction. Due to many influencing factors, such as hybridization, reticulate evolution, horizontal transfer, interaction among loci, recombination, and coevolution, the genetic tree based on a single genetic DNA fragment is not necessarily consistent with the real evolutionary tree of the species. In addition, the chloroplast is maternally inherited, so the whole genome of chloroplast is likely to represent the evolutionary relationship of the species only in the absence of hybridization or chloroplast capture events. Based on the above characteristics, chloroplast genome analysis plays an important role in plant phylogeny research.

1.3.2 Current phylogenetic studies in the *Aquilegia* genus

In the past, researchers paid increased attention to the *Aquilegia* genus, as it was a new model system for evolutionary studies due to its phylogenetic mid-position between eudicots and monocots (Kramer and Hodges, 2010). Former studies in *Aquilegia* mostly focused on unusual floral organs, such as petaloid sepals, the staminodium, and petals with a nectar spur, which stimulated the study of phylogeny and evolution in this genus (Kramer et al., 2007; Voelckel et al., 2010). For instance, floral isolation evolved between *Aquilegia formosa* and *Aquilegia pubescens*, indicating that variation in nectar spurs would influence reproductive isolation (Fulton and Hodges, 1999). Later studies used molecular data to provide a fundamental phylogenetic relationships in *Aquilegia*.

With the technological advancement, more meticulous and in-depth studies have been undertaken. For instance, 1) through an amplified fragment length polymorphism (AFLP) phylogeny, researchers inferred several independent losses of floral anthocyanins in the North American *Aquilegia* clade. This study also provided a first phylogenetic reconstruction of the genus, and the indication of its Asian origin (Whittall et al., 2006). 2) Based on nuclear internal transcribed spacer (ITS) sequences and *trnK-matK* and *trnS-G* regions of chloroplast DNA, phylogenetic studies were poorly solved among 32 *Aquilegia* taxa from North America, Europe and Asia, indicating that few sequence data were unable to resolve infra-generic relationships as a result of low nucleotide variation among taxa (Bastida et al., 2010). 3) The phylogenetic analysis of *Aquilegia* by 34 single copy genes have indicated that the genetic basis plays an important role in evolution of flower color (Hodges and Derieg, 2009). This study conjugated genomic information with genetic basis of adaptation to aid our understanding of evolutionary drivers in the *Aquilegia* genus. 4) Analyses of 21 non-coding plastid sequences from 84 *Aquilegia* taxa have demonstrated a preliminary assessment of evolutionary relationships among North American, Asian and European taxa in this genus (Fior et al., 2013). Although this work had divided *Aquilegia* into five group, the large European clade was not solved completely because of lack of sufficient phylogenetic information. Taken together, all these researches demonstrated that sufficiently large amounts of the genetic and genomic information from *Aquilegia* are needed for phylogenetic and evolutionary studies in this genus.

1.3.3 Current phylogenetic studies in the Brassicaceae family

The traditional classification systems of Brassicaceae are generally focused on morphological characters including fruit shape, position of cotyledons, trichome types, nectar-gland morphology, etc. For instance, Judd defined that Brassicaceae were nested within the paraphyletic Capparaceae (Judd et al., 1994). Stevens classified the order

Brassicales extending the order Capparales to comprise 17 families, 398 genera, and roughly 4,450 species (Stevens, 2001). However, molecular studies have recurrently supported that Brassicaceae is sister to Cleomaceae and both are sister to Capparaceae (Hall et al., 2002a; Hall et al., 2004; Schranz and Mitchell-Olds, 2006). Therefore, these three families are currently recognized in Brassicales.

Using only morphological character information, it is, however, very difficult to understand and clarify the evolutionary relationships among taxa of Brassicaceae because of their extensive convergence and parallel evolution (Franzke et al., 2011). A great number of molecular phylogenetic analyses have recently resolved many of the different problems caused by the incongruence between molecular morphological studies and molecular biology studies, and have identified many monophyletic species in Brassicaceae that are supported by both morphological and molecular researches. Now, in fact, most phylogeny studies in the Brassicaceae family are using morphological data and molecular data (Marhold et al., 2002; Scheen et al., 2002; Perný et al., 2005), but more and more species and molecular data are needed for in-depth analyses.

As to Brassicaceae, the tribal classification systems have been summarized in various reviews. For instance, the new classification system of tribes in Brassicaceae has been summarized in 2006 by Beilstein and his colleagues, who have divided the core part of the family into three large monophyletic groups and other small monophyletic groups based on the results of the chloroplast *ndhF* gene (Beilstein et al., 2006). Lineage I contains Camelineae, Boechereae, Halimolobeae, Cardamineae, Descurainieae, Physarieae, Lepidideae and Smelowskieae; Lineage II includes Brassiceae, Isatideae, Schizopetaleae and Sisymbrieae; Lineage III comprises Anchonieae, Chorisporeae, Euclidieae, and Hesperideae; other small monophyletic groups represent one tribe, such as Alysseae, Arabideae, Eutremeae, Heliophyileae, Cochlearieae, Iberideae and Thlaspideae. This significant conclusion is also supported by other phylogenetic analyses of different data (nuclear gene sequences, mitochondrial and chloroplast gene sequences), such as ITS, *nad4* first intron, *ndhF* and *phyA, adh, chs, trnL-F* and *matK*, etc. (Bailey et al., 2006; Al-Shehbaz et al., 2006; Franzke et al., 2009; Beilstein et al., 2008; Couvreur et al., 2010).

1.4 Molecular evolutionary studies of plastomes

The field of molecular evolution can be grouped into two types of investigations: studies of phylogeny and studies of the molecular evolutionary process. It is generally known that molecular evolution is the process of changes in the sequences (DNA, RNA and proteins) across generations. Major topics in molecular evolution concern the rates and impacts of single nucleotide variation, relative importance of neutral drift and natural selection, origin of new genes, the heritability of complex traits, the genetic basis of

speciation, development of evolution, and the effects of evolutionary forces on genomes and traits.

1.4.1 The driving forces of evolution

There are three perspectives for molecular evolution according to the relative importance assigned to the various forces of evolution (Graur and Li, 2000):

1) The first one, selectionist hypothesis, emphasizes that selection is the driving force of molecular evolution. Although many mutations are neutral, researchers attribute changes in the frequencies of neutral alleles to linkage disequilibrium with other loci that are under selection, rather than to random genetic drift (Hahn, 2007). Codon usage bias refers to differences in the frequency of occurrence of synonymous codons in coding DNA. The overabundance in the number of codons allows many amino acids to be encoded by more than one codon. The genetic codes of different organisms are often biased towards using one of the several codons that encode the same amino acid over the others.

2) The second one, neutralist hypothesis, pays attention to the importance of mutation, purifying selection, and random genetic drift. The neutral theory of molecular evolution proposes that harmful mutations are quickly removed, and other mutations in DNA are not important to function or fitness (Kimura, 1968). The fate of neutral mutations is governed by genetic drift, and contributes to both nucleotide polymorphism and fixed differences between species (Nachman, 2006).

3) The third one, mutationist hypothesis, focuses on random drift and bias in mutation patterns (Nei, 2005). Sueoka has proposed that the variation in GC content is a consequence of the GC mutational pressure, not the result of positive selection (Sueoka, 1964). A prominent feature of chloroplast DNA (cpDNA) is its low guanine and cytosine (GC) content. Indeed, all of the 1,234 completely-sequenced chloroplast genomes available at the NCBI as of May 2017 have a GC content between 28.36 and 43.29% (average = 37.39%, SD=1.46%). It is known that the strong AT (adenine and thymine) bias is reflected in codon usage, where an A or T is preferred in the third position of synonymous codons (Shimda and Sugiuro, 1991). Furthermore, researchers have found that the patterns of codon usage greatly differ between monocot and dicot species (Liu and Xue, 2005). On the other hand, strong evidence from nematode nuclear genomes shows that GC content influences both codon usage and amino acid composition and that GC content is probably driven by directional mutation pressure (Mitreva et al., 2006).

1.4.2 Different forces in molecular evolution

It is commonly known that the forces from molecular and population genetics influence the content and structure of the chloroplast genome. Novel genetic variants arise through mutation and frequency changes in populations due to genetic drift or natural selection. Most mutations are point mutations resulting from single nucleotide polymorphisms which modify single bases of the DNA sequence. And other types of mutations are duplication, insertion, deletion, inversion, and translocation. Actually, most organisms display a strong bias in the types of mutations which affect the probability of occurrence of GC content. Using the mutation rate per generation and the number of nucleotide difference between two sequences, divergence times are estimated effectively, that is under the assumption of a molecular clock. Genetic drift is the change of allele frequency from one generation to the next due to stochastic effects of random sampling in finite populations. Some existing variants have no effect on fitness and may increase or decrease in frequency simply due to chance.

Selection occurs when individuals have different fitness, for example, greater ability to survive or reproduce, favoring the inheritance of their genetic background in subsequent generations, thereby increasing the instances of the underlying genetic variants in a population. Selection includes natural selection, artificial selection, or sexual selection. Natural selection is the differential survival and reproduction of individuals because of difference in phenotype. It is a key mechanism of evolution, and the changes can happen in basic any heritable traits of populations over time. Besides, natural selection acts at different levels of organization, such as genes, cells, individual organisms, groups of organisms and species (Gould, 1998).

1.4.3 Evolutionary rate

Detecting adaptive evolution at the genetic level helps to understand the structural and functional variation of genes and the evolutionary history of organisms (Nei and Kumar, 2000). Previous studies have established that rates of molecular evolution vary among sites (e.g., amino acid residues) and among proteins (e.g., cytochrome c versus hemoglobin) (Ossowski et al., 2010; Drummond and Wilke, 2008). Evolutionary rates of nucleotide are even more complex, as the genetic code is intrinsically redundant. There are two main aspects influencing the variability of evolutionary rates among nucleotide sites (Gaut et al., 2011). First, natural selection acts differentially among sites. Following mutation, positive selection increases the frequency of alleles that confer a fitness advantage to the individuals bearing them, thus temporarily increasing the genetic diversity of the population, while negative selection removes those alleles that are deleterious and decreases genetic diversity. Codon-based models of molecular evolution

are able to infer signatures of selection from alignments of homologous sequences by estimating the relative rates of synonymous (dS) and non-synonymous substitutions (dN). Non-synonymous sites typically evolve more slowly than synonymous sites, presumably because amino acid replacements are functionally constrained. The ratio between dN and dS (also commonly called omega, ω) allows to estimate the type and intensity of selection acting on different codons. Under neutral evolution, the average omega is expected to be close to 1, while under positive or purifying selection it is significantly higher or lower than 1, respectively. When sufficient sequences are aligned, the stochastic effects of omega sampling become negligible and it appears clearly that different codon positions evolve according to very different selective regimes. Second, also the underlying mutation process varies among sites. For example, methylated cytosines, deaminate spontaneously. Deamination leads to high mutation rates and the preferential replacement of cytosine (C) with thymine (T). While most attention has been traditionally devoted to the elucidation of patterns of positive or purifying selection, the absence or reduction in efficiency of natural selection can allow the onset of a relative increase of dN over dS, known as relaxed selection (Wertheim et al., 2015). From a methodological point of view, when studying single genes relaxed selection can be easily confused with positive (directional) selection, so that its relevance to drive organismal evolution has been underestimated for a long time. However, due to the gene-wide or genome-wide reduction in the efficiency or intensity of purifying and positive selection, relaxed selection can play a very relevant evolutionary role in the exploration of a wider subset of the phenotypic space and thus foster evolutionary innovation.

Several factors can in principle affect the estimation of omega. One of the factors most commonly disregarded is recombination. By mixing up the sequence of different genes or alleles, recombination can obscure the real evolutionary patterns and mislead inference of positive selection. In the case of chloroplast genes, recombination is not expected to be a major disturbance factor, as the maternal inheritance of the plastome as a single locus greatly reduces the possibility of chloroplast genes to recombine. On the other hand, RNA editing is a molecular process through which some cells make discrete changes to specific nucleotide sequences within a RNA molecule after it has been generated by RNA polymerase. As such, the inference in DNA sequences for what is a synonymous versus a non-synonymous substitution can easily be misled. RNA editing occurs in the cell nucleus and cytosol, as well as within mitochondria and plastids (Danecek et al., 2012; Takenaka et al., 2014; Shikanai, 2015). The diversity of RNA editing phenomena includes nucleobase modifications such as cytidine (C) to uridine (U) and adenosine (A) to inosine (I) deaminations, but C-to-U editing often occurs in the mitochondrion and chloroplast RNA of flowering plants. It is generally known that RNA editing is essential for the normal functioning of the plant's translation and respiration activity (Price and Gray, 1998).

1.5 The aim of this work

Plant chloroplast genomes are highly conserved in size, gene content and structure, however, their molecular evolutionary drivers and relevance for environmental adaptation remain poorly understood. Although the phylogeny of flowering plants could help to better understand plastome evolution in general, complete plastomes of plants have been sequenced only for some species so far, and the taxonomic and geographic sampling is sparse. In this study, to enhance understanding of the phylogenetic relationships among some important taxa and the specific mechanisms of evolution both at the molecular and gene levels, different levels were analyzed as following:

1) At genus level: as the large European *Aquilegia* clade was not resolved completely because of lack of sufficient phylogenetic information, 34 *Aquilegia* taxa were sequenced and 66 CDS were obtained. Several aspects were tested to ascertain whether: 1) the overall GC content and GC content in the third position of codon has correlation with amino acid polymorphisms and codon usage frequencies in chloroplast genes of *Aquilegia*; 2) these coding sequences characteristics would explore the connection between diversification and evolution in this genus; 3) the phylogenetic relationships of 34 *Aquilegia* taxa would provide novel insights into the patterns of radiation of the genus in Europe.

2) At family level: as the tribal classification of Brassicaceae was not fully resolved, 78 sampled species and 17 references (NCBI) in the Brassicaceae family were analyzed based on 71 protein coding genes. This part of the study tested whether: 1) whole-gene positive selection would differentially affect different gene categories; 2) it would be possible to define more precisely the classification of Brassicaceae taxa at the molecular level of chloroplast DNA; 3) compared to mixtures of coding and non-coding data, the tribal classification based on a large amount of protein coding sequences would similar.

3) Across families: the protein coding genes of 11 families were analyzed to reveal the pattern of chloroplast DNA evolution across a subset of families in angiosperms at the molecular level. As chloroplasts play a crucial role in sustaining life on Earth, this part of the work was aimed at investigating whether and, in case, to what extent chloroplast genomes bears signatures that hint at involvement of chloroplast function into the selective processes leading to plant adaptation and differentiation. In particular, by using each family as a proxy of independent evolutionary replicate of plant species adaptation and colonization of the alpine environment, this part of the work tested whether: 1) recurrent patterns of whole-gene positive selection would affect chloroplast CDS in the different families; 2) positive selection patterns would be randomly distributed or not along the branches of each family phylogenetic tree; 3) elevational adaptation of species left any differential signature of positive selection in high and low altitude plastomes; 4) specific positions of the genes showed particularly evident signs of positive selection,

which could hint at the mechanisms underlying the observed omega increase. From a methodological point of view, these results further indicate that the use of gene trees and of species trees provide consisting results for evolutionary analyses of chloroplast CDS and that effect of RNA editing has a detectable but marginal effect on the assessment of plastome evolution.

2 Materials and Methods

2.1 Materials

2.1.1 The *Aquilegia* genus

In total, 34 species of *Aquilegia* were collected, of which 28 species were from Europe and 6 species were outgroups from Asia (Table 2.1)

Table 2.1: Accessions sampled of the *Aquilegia* genus

Taxonomy ID	Sample	Locality	Reference
218850	*A. alpina*	Europe	Fior et al., 2013
560536	*A. atrata*	Europe	Fior et al., 2013
560537	*A. aurea*	Europe	Fior et al., 2013
560538	*A. barbaricina*	Europe	Fior et al., 2013
560539	*A. bernardii*	Europe	Fior et al., 2013
1291431	*A. bertolonii*	Europe	Fior et al., 2013
1291432	*A. blecicii*	Europe	Fior et al., 2013
1277885	*A. buergeriana*	Asia	Fior et al., 2013
1291433	*A. dinarica*	Europe	Fior et al., 2013
1291434	*A. dumeticola*	Europe	Conti et al., 2005
560540	*A. einseleana*	Europe	Fior et al., 2013
560543	*A. glandulosa*	Asia	Fior et al., 2013
1291436	*A. grata*	Europe	Fior et al., 2013
1291437	*A. iulia*	Europe	Nardi, 2011
1291440	*A. kitaibelii*	Europe	Fior et al., 2013
1291442	*A. litardierei*	Europe	Fior et al., 2013
1291444	*A. magellensis*	Europe	Conti et al., 2005
560545	*A. nigricans*	Europe	Fior et al., 2013
560546	*A. nugorensis*	Europe	Fior et al., 2013
1291447	*A. nuragica*	Europe	Fior et al., 2013
349364	*A. olympica*	Asia	Fior et al., 2013
1291448	*A. ottonis* subsp. *amaliae*	Europe	Fior et al., 2013
1291449	*A. ottonis* subsp. *speluncarum*	Europe	Fior et al., 2013
1291451	*A. oxysepala* var. *oxysepala*	Asia	Fior et al., 2013
1291452	*A. pancicii*	Europe	Fior et al., 2013
349363	*A. pyrenaica*	Europe	Fior et al., 2013
1291455	*A. reuterii*	Europe	Fior et al., 2013
560549	*A. sibirica*	Asia	Fior et al., 2013
560551	*A. thalictrifolia*	Europe	Fior et al., 2013
1291457	*A. transsilvanica*	Europe	Fior et al., 2013
1291458	*A. vestinae*	Europe	Conti et al., 2005
1506457	*A. viridiflora*	Asia	Fior et al., 2013
560553	*A. viscosa*	Europe	Fior et al., 2013
1291459	*A. vulgaris*	Europe	Fior et al., 2013

2.1.2 The Brassicaceae family

About 78 species of Brassicaceae were collected in Italy (Table 2.2), and 17 reference chloroplast genomes downloaded from NCBI database were listed in Table 2.3.

Table 2.2: Accessions sampled of the Brassicaceae family

Taxonomy ID	Sample	Reference
228874	*Berteroa incana*	Bailey et al., 2007
169068	*Alyssum alissoides*	Huang et al., 2015
358668	*Fibigia clypeata*	Koch et al., 2007
369019	*Matthiola fruticulosa*	Warwick et al., 2006
358661	*Bunias orientalis*	Beilstein et al., 2006
161951	*Draba verna*	Jordon-Thaden et al., 2010
87303	*Draba dubia*	Schwienbacher et al.,2011
50452	*Arabis alpina*	Beilstein et al., 2006
78191	*Arabis hirsuta Aggregate*	Warwick et al., 2006
571360	*Arabis nova*	Mutlu, 2004
81975	*Arabis soyeri* subsp. *subcoriacea*	Warwick et al., 2006
81982	*Arabis turrita*	Koch et al., 2001
648807	*Boechera gracilipes*	Dorn, 2003
359876	*Phoenicaulis cheiranthoides*	Beilstein et al., 2006
115933	*Polyctenium fremontii*	Beilstein et al., 2006
264416	*Diplotaxis tenuifolia*	Martínez-Sánchez et al., 2007
1035077	*Brassica repanda* subsp. *baldensis*	Prosser and Bertolli, 2007
71354	*Hirschfeldia incana*	Beilstein et al., 2006
71324	*Camelina microcarpa*	Beilstein et al., 2006
264402	*Capsella grandiflora*	Koch and Mummenhoff, 2006
-	*Erysimum aurantiacum*	Gilardelli et al., 2013
761866	*Erysimum rhaeticum*	Müller et al., 2006
1370092	*Erysimum sylvestre*	Regvar et al., 2006
1370093	*Erysimum virgatum*	Warwick et al., 2006
98023	*Neslia paniculata*	Bailey et al., 2006
65952	*Rorippa sylvestris*	Huang et al., 2015
50463	*Cardamine hirsuta*	Lihová et al., 2006
352360	*Cardamine alpina*	Lihová et al., 2009
50462	*Cardamine flexuosa*	Lihová et al.,2006
157082	*Rorippa austriaca*	Bleeker and Matthies, 2005
416611	*Dentaria enneaphyllos*	Sebastia et al., 2005
82288	*Dentaria pentaphyllos*	Surina, 2002
70807	*Leavenworthia uniflora*	Charlesworth et al., 1998
341031	*Leavenworthia exigua*	Charlesworth et al., 1998
270110	*Cochlearia officinalis*	de Vos et al., 2013
358665	*Descurainia bourgaeana*	Koch and Mummenhoff, 2006
89411	*Descurainia sophia*	Koch and Mummenhoff, 2006
153321	*Hornungia petraea*	Kluth and Bruelheide, 2005
153458	*Hutchinsia alpina*	Urbanska, 1997
190879	*Hutchinsia brevicaulis*	Ančev, 2007

190882	*Hymenolobus pauciflorus*	Fonderflick et al., 2010
369011	*Malcolmia littorea*	Gratani et al., 2011
369026	*Morettia philaeana*	Khalik et al., 2002
98038	*Thellungiella halophila*	Koch and Mummenhoff, 2006
167691	*Halimolobos pubens*	Bailey et al., 2007
263653	*Heliophila coronopifolia*	Saito et al., 2011
264418	*Hesperis matronalis*	Beilstein et al., 2006
190884	*Iberis amara*	Bailey et al., 2007
161756	*Isatis tinctoria*	Beilstein et al., 2006
65351	*Lepidium campestre*	Huang et al., 2015
153317	*Cardaria draba*	Kiemnec and McInnis, 2002
473028	*Noccaea praecox*	Regvar et al., 2013
1230357	*Noccaea rotundifolium*	Warwick et al., 2006
-	*Lesquerella montana*	Al-Shehbaz and O'Kane, 2002
-	*Nerisyrenia camporum*	Hall et al., 2002b
72662	*Stanleya pinnata*	Beilstein et al., 2006
359899	*Thelypodium laciniatum*	Beilstein et al., 2006
664029	*Ochthodium aegyptiacum*	Khalik et al., 2002
203582	*Sisymbrium officinale*	Huang et al., 2015
98035	*Smelowskia calycina*	Beilstein et al., 2006
126278	*Thlaspi perfoliatum*	Guimarães et al., 2009
-	*Peltaria angustifolia*	Aghaei et al., 2013
264427	*Biscutella laevigata*	Parisod and Besnard, 2007
380183	*Biscutella prealpina*	Tremetsberger et al., 2002
71322	*Calepina irregularis*	Huang et al., 2015
169074	*Kernera saxatilis*	Warwick et al., 2006
153659	*Lunaria annua*	Beilstein et al., 2006
228870	*Cleome spinosa*	Marshall et al., 2007
860697	*Cleome hirta*	Marshall et al., 2007
457767	*Alyssum dasycarpum*	Turgay et al., 2012
87302	*Draba aizoides*	Schwienbacher et al.,2011
63678	*Turritis glabra*	Beilstein et al., 2006
82288	*Cardamine pentaphyllos*	Sweeney and Price, 2000
228783	*Cardamine asarifolia*	Lihová et al., 2004
352363	*Cardamine trifolia*	Philippe and Ochyra, 2004
50465	*Cardamine pratensis*	Lihová et al., 2004
202778	*Aethionema saxatile*	Beilstein et al., 2006
81970	*Arabidopsis halleri*	Bailey et al., 2007

Table 2.3: 17 references used in Brassicaceae analyses

Species name	Accession_Genebank	Reference
Aethionema cordifolium	NC_009265	Hosouchi et al., 2007
Aethionema grandiflorum	NC_009266	Hosouchi et al., 2007
Arabidopsis thaliana	NC_000932	Sato et al., 1999
Arabis hirsuta	NC_009268	Hosouchi et al., 2007
Barbarea verna	NC_009269	Hosouchi et al., 2007
Brassica napus	NC_016734	Hu et al., 2011
Capsella bursa-pastoris	NC_009270	Hosouchi et al., 2007
Crucihimalaya wallichii	NC_009271	Hosouchi et al., 2007
Draba nemorosa	NC_009272	Hosouchi et al., 2007
Lepidium virginicum	NC_009273	Hosouchi et al., 2007
Lobularia maritima	NC_009274	Hosouchi et al., 2007
Nasturtium officinale	NC_009275	Hosouchi et al., 2007
Olimarabidopsis pumila	NC_009267	Hosouchi et al., 2007
Pachycladon cheesemanii	NC_021102	Becker et al., 2013
Pachycladon enysii	NC_018565	Becker et al., 2013
Cardamine resedifolia	NC_026446	Hu et al., 2015
Cardamine impatiens	NC_026445	Hu et al., 2015

2.1.3 Other important angiosperms species

A total of 1037 species were collected in the Trentino-South Tyrol region, and detailed information could be found in Table 3.7.

2.2 Bioinformatics methods

2.2.1 DNA extraction and sequencing

In total, approximately 1,150 species of angiosperms were collected. At first, the DNA of the fresh leaves of each plant sample was extracted using the DNeasy Plant Mini kit (Qiagen GmbH, Germany) according to the manufacturer's instructions. Then, agarose gel electrophoresis was applied to check DNA quality. Last, total DNA samples were sequenced on Illumina HiSeq2000 sequencer. The above phases of the experiment work were carried out by laboratory staff (Dr. Mingai Li and Enrico Barbaro, Research and Innovation Center, Fondazione Edmund Mach, Italy). The following analyses were carried out according to the pipeline summarized in Figure 2.1.

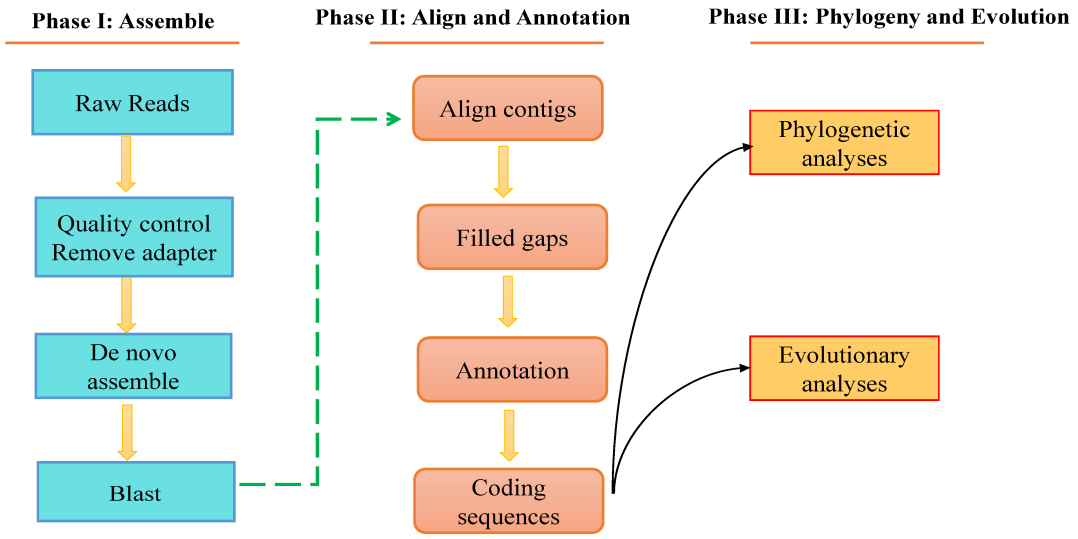

Figure 2.1: The pipeline of bioinformatics analyses

2.2.2 Chloroplast genome assembly

Subsequent to Illumina sequencing, phiX174 contaminants (resulting from spiking internal controls used for sequencing) were removed from the pool of raw reads through reads mapping to the phage genome with the bowtie program using the parameters –t –p 10 –un –S (Langmead et al., 2009). Afterwards, FastQC v0.11.2 (Andrews, 2010) was applied to check the quality of reads and provide a first visual inspection of their quality. According to the results of FastQC, the TruSeq adapters were removed by Trimmomatic v0.32 (Bolger et al., 2014) using the following settings: PE -phred33 ILLUMINACLIP: TruSeq3-PE-2.fa:2:30:10 LEADING: 3 TRAILING: 3 SLIDINGWINDOW: 4:15 MINLEN: 100, and the reads were trimmed by FASTX-Toolkit v0.0.14 (Gordon and Hannon, 2012) with parameters –q 20 –p 90 –Q 33. Then, FastQC was used once again to carry out the final reads quality check. To obtain the *de novo* plastome assembly, Velvet v1.2.10 (Zerbino and Birney, 2008) was utilized to assemble reads. In this step, the K-mers were evaluated from 35 to 75 in increments of 4 for a representative subset of species. Based on the results of this preliminary analysis, K-mer = 63 was used for assembly of all the species in velveth. In velvetg, the parameters -cov_cutoff auto -ins_length 300 -exp_cov auto -min_contig_lgth 100 were set for assembly. Subsequently, calculating the number of contigs, total base pairs, N50 and N90 would receive a basis for assessing the quality of assembly. NCBI-blast-2.2.31 (Camacho et al., 2009) was used to compare contigs with plastids genomes, and then the best match reference for each species was chosen with parameters -evalue $1e^{-5}$ -max_target_seqs 2.

2.2.3 Gene annotation and alignment

2.2.3.1 The *Aquilegia* genus

According to the reference *Thalictrum coreanum* plastid (Park et al., 2015), ABACAS (Algorithm Based Automatic Contiguation of Assembled Sequences) (Assefa et al., 2009) program with parameters -p nucmer -c and MUMmer (Kurtz et al., 2004) program with nucmer and mummerplot packages were performed to contiguate plastomes. In order to fill the gaps of these pseudomolecules, Gapcloser v1.10 was employed using default parameters (http://soap.genomics.org.cn/index.html). At last, MAUVE v2.3.1 (Darling et al., 2004) was applied to visualize the rearrangement and inversion between samples and reference.

Based on the gap-filled plastomes of *Aquilegia*, two different approaches, DOGMA (Dual Organellar GenoMe Annotator) and a perl script which locally automates the annotation process, were applied for annotation. DOGMA is a website for annotating plant chloroplast and animal mitochondrial genomes, and 60% as identity cutoff for protein coding genes and 1e-5 as e-value cutoff were used in this study (Wyman et al., 2004). On the other hand, the perl script compares the plastomes with reference sequence and locates homologous regions on the plastomes for each gene sequence feature (http://genomics-pubs.princeton.edu/prv/resources/scripts/migrate_annotations.pl).
Finally, these two methods were integrated and manually edited the start/stop codons and intron regions in DOGMA program. Extracted chloroplast CDS per species from the DOGMA website were used for phylogenetic and evolutionary analyses.

Then the MACSE v1.01 (Ranwez et al., 2011) was applied to align each coding nucleic acid sequence with respect to their amino acid translation, and stop codons for each alignments were deleted by the ReplaceStopsWithGaps.pl script (https://gist.github.com/josephhughes/1167776). After that, these alignments were concatenated to generate a dataset which consisted of 34 species and each individual had 66 CDS by means of custom perl scripts. Although this dataset provided a first assessment of the phylogenetic relationships among Eurasian *Aquilegia* taxa, the whole plastome sequences (only partially complete for some taxa) were employed as a complementary approach for the inference of phylogenetic relationships among species. The plastome pseudomolecules which were obtained from the MUMmer and ABACAS programs were aligned with the Kalign program (Lassmann and Sonnhammer, 2005). After manually deleting the regions containing Ns, poorly aligned regions were further eliminated by application of the Gblocks v0.91b program (Talavera and Castresana, 2007).

2.2.3.2 The Brassicaceae family

The contigs were arranged by the ABACAS (Assefa et al., 2009) and MUMmer programs (Kurtz et al., 2004) with nucmer and show-tilling packages to contiguate plastomes. Then Gapcloser from the SOAP software packages was applied to fill gaps for the plastomes using default parameters (Li et al., 2008). Blastn (Camacho et al., 2009) was carried out to compare gap-filled plastomes with reference chloroplast genomes, and the relevant inverted repeat (IR) region was observed and the lacking IR region was filled up through the Ugene toolbox (Okonechnikov et al., 2012) and custom perl scripts. The number of Ns and the total length of plastomes were calculated. Afterwards, two methods were applied to annotate the whole plastome sequences. One approach was DOGMA (Wyman et al., 2004), setting 60% as identity cutoff for protein coding genes and 1e-5 as e-value cutoff. The other approach was to carry out the annotation by a previously published perl script (Szpara et al., 2011). Finally, these two methods were integrated and manually edited the start/stop codons and intron regions in the DOGMA program, then the CDS sequences per species from the DOGMA website were extracted. Last MACSE v1.01 (Ranwez et al., 2011) was applied to align each coding nucleic acid sequence with respect to their amino acid translation, and stop codons were deleted by the custom perl ReplaceStopsWithGaps.pl script.

2.2.3.3 Other important angiosperms species

After assembling, all mitochondrial genomes of plants were downloaded from the NCBI database, which included 238 species and 7206 protein sequences. Mitochondrial CDS were selected by running blastx homology searches (NCBI-blast-2.2.31) with contigs retaining all putative genes with identity ≥ 80% and length ≥ 80% of the reference protein sequences. A second filtering step was performed using a dataset of curated mitochondrial genes from 15 well annotated mitochondrial genomes, representing 7 widely divergent angiosperm species (Kubo and Newton, 2008). The number of ribosomal subunits, the large subunit (LSU) and small subunit (SSU), were further calculated. The LSU and SSU sequences (also 5.8S) of *Viridiplantae* from SILVA rRNA database (https://www.arb-silva.de/) were downloaded. Then rRNA genes were selected by running blastn homology searches retaining all putative genes with identity ≥ 70% and qcovs ≥ 60%. The genes which also resulted to hit plastids and mitochondria by blastn were removed. Last these sequences were aligned by the MAFFT program (Katoh et al., 2005) and low homology regions were removed using the Gblocks program (Castresana, 2000).

The contigs were contiguated as pseudomolecules based on reference genomes using MUMmer with nucmer, show-tilling (with -a –v 50), and mummer plot (with --large --

layout --color --png) packages (Kurtz et al., 2004). Additionally, the cleaned reads were used to fill gaps in those pseudomolecules by the Gapcloser program (http://soap.genomics.org.cn/index.html) with default parameters. Then they were annotated using a perl script (annotations.pl: http://genomics-pubs.princeton.edu/prv/resources/scripts/migrate_annotations.pl). Extracted as described above, the CDS sequences per species were selected according to the annotation results, and the species which had no less than 20 protein coding sequences were chosen. Then they were organized in the different families according to their taxonomic classification in the APG IV. For each family, MACSE v1.01 (Ranwez et al., 2011) was applied to align each coding nucleic acid sequences with respect to their amino acid translation with parameters -prog alignSequences -gc_def 11, and stop codons and gaps were deleted. Then, the sequences whose alignments length was no less than 80% of the reference gene were kept. Finally, BioEdit v7.2.3 was used to visually check all the alignments before phylogeny and molecular evolution analyses.

2.2.4 Phylogenetic analyses

2.2.4.1 Phylogenetic analyses in the *Aquilegia* genus

Based on the above mentioned two datasets, several different methods were used to construct phylogenetic trees for studying the evolutionary relationships among Eurasian *Aquilegia*. Firstly, jModelTest (v2.1.4) program with parameters base frequencies (+F) and rate variation (+I, +G) was employed to carry out statistical selection of best-fit models of nucleotide substitution by the Akaike Information Criterion (AIC) (Akaike, 1973; Darriba et al., 2012). Secondly, CodonPhyML v1.00 (Gil et al., 2013) analysis selected GY (Goldman and Yang, 1994) as substitution model and F3X4 as frequency model. Thirdly, according to the result of jModelTest, maximum likelihood (ML) analysis was utilized by PhyML v3.1 (Guindon et al., 2009) with parameters custom model and p-invar 0.98, and maximum parsimony (MP) analysis was performed by PHYLIP v3.695 with parameters 5 random seed and majority rule (consensus type). Both of ML and MP analyses were subjected to bootstrap resampling (100 replicates) to estimate robustness. Furthermore, bayesian inference (BI) was performed by Mrbayes v3.2.2 (Ronquist and Huelsenbeck, 2003) with settings: lset nst=6 rates=invgamma Ngammacat=6 mcmcp ngen=2000000 diagnfreq=1000 samplefreq=1000 printfreq=1000 nchains=4. Finally, the phylogenetic trees were displayed through Treegraph2, which combined and visualized evidence from different trees (Stöver and Müller, 2010).

2.2.4.2 Phylogenetic analyses in the Brassicaceae family

Based on the 71 chloroplast CDS of 95 species in the Brassicaceae family, several different methods were used to reconstruct phylogenetic trees. Firstly, jModelTest v2.1.4 program was employed to carry out statistical selection of best-fit models of nucleotide substitution by AIC, and ProtTest program was applied for selecting the best amino acid model (Darriba et al., 2011). Secondly, codonPhyML obtained a phylogenetic relationships based on codon by using DNA sequences (Gil et al., 2013). ML trees were also constructed through the PhyML program (Guindon et al., 2009) with general time reversible (GTR+I+G) model for nucleotide substitution and JTT model for amino acid substitution. MP trees were performed by using PAUP* 4.10 (Swofford, 2003). ML and MP analyses were subjected to bootstrap resampling (100 replicates) to estimate robustness. BI trees were performed by using MrBayes (Ronquist et al., 2003). RAxML web servers were used to calculated aLRT, a rapid bootstrap algorithm (Stamatakis et al., 2008). Finally, the phylogenetic trees were displayed through Treegraph2, which combined and visualized evidence from different trees (Stöver and Müller, 2010).

2.2.4.3 Other important angiosperms species

Based on gene alignments of each family, CodonPhyML v1.00 (Gil et al., 2013) was used to construct phylogenetic trees for each gene studying the relationships among plant species by selecting GY as substitution model and F3X4 as frequency model. To save computational time, SH-like branch value supports were calculated, because approximate likelihood ratio test (aLRT) is a fast and accurate method to infer branch support (Anisimova and Gascuel, 2006). Given the partial overlap in gene sampling across species in this work, the species trees for each family were reconstructed using a supertree approach. The CLANN program was used to construct super-trees performing 100 bootstrap iterations to test the robustness of the inferred topology (Creevey and McInerney, 2005). For each supertree, clustering of congeneric species was visually checked to assess the topological agreement with the accepted taxonomy of the taxa analyzed. Taxa with incongruent placement with respect to genus delimitation were individually checked for synonymy and/or recent renaming. Only a few cases that could not be resolved were excluded from subsequent analyses, due to possible mislabeling.

2.2.5 Evolutionary analyses

2.2.5.1 The GC content and codon usage analyses in the *Aquilegia* genus

By alignment of the 66 CDS through MACSE with default genetic code (gc_def 11), nucleic acids and the corresponding amino acids alignments were simultaneously

obtained. Through the BioEdit v7.0.0 (Hall, 1999) program, the amino acid polymorphisms in these alignments were observed, as the occurrence of amino acid polymorphisms was derived from nucleotide site changes (Rand and Kann, 1996).

It is well accepted that codon usage is a statistical property of the protein encoding regions in DNA sequences, and the degeneracy of the genetic code implies that one amino acid can be encoded by several codons (Guilloux and Jestin, 2012). The frequencies of codon usage in 66 CDS alignments were counted using the online tool "codon usage" (http://www.bioinformatics.org/sms2/codon_usage.html), in which the genetic code was set as 11 (Bacterial and Plant Plastid). Besides, codon usage frequency in chloroplast genes of the other three published *Ranunculaceae* taxa (*Megaleranthis saniculifolia*, *Ranunculus macranthus*, *Thalictrum coreanum*) and two angiosperms (*Arabidopsis thaliana* and *Oryza sativa Japonica*) (Kim et al., 2009; Raubeson et al., 2007; Park et al., 2015; Sato et al., 1999; Morton and Clegg, 1993) were calculated and compared with *Aquilegia* samples.

2.2.5.2 Positive selection analysis in the Brassicaceae family

To reveal the pattern of molecular evolution in Brassicaceae, Selecton program was applied to identify positive selection on 71 alignment sequences in 95 Brassicaceae taxa. Selecton is an evolutionary codon model which enables calculating ω (omega) at each codon site using a maximum-likelihood approach (Stern et al., 2007). In this work, M8 model (Yang et al., 2000) and M8a (Swanson et al., 2003) were chosen as positive-selection enabling model and null model to analyze these 71 protein coding genes, respectively.

2.2.5.3 Natural selection and RNA editing analyses in the other important angiosperms species

To identify the role of selection on the evolution of chloroplast genes across families, coding sequences of alignments were analyzed by using the HyPhy packages (Pond et al., 2005): gene-wide selection was evaluated with BUSTED (Murrell et al., 2015); branch-level selection with aBSREL (Smith et al., 2015); site-level selection with FUBAR and MEME (Murrell et al., 2013; Murrell et al., 2012a); relaxed or intensified selection associated to elevational preference with RELAX (Wertheim et al., 2015). The "Universal" genetic code was used in all analyses. The option to test all the branches for the packages based on branch-level analysis was chosen. The GTR model was applied in the detection of sites under positive selection (MEME package), and only estimating the dN/dS score. Additionally, only tests with probability lower than 0.05 were considered significant and classified under positive selection. For the final presentation of the data,

the false discovery rate (FDR) correction for multiple tests was applied to account for the comparison between different genes and families as a single family of related tests. Strictly speaking, as all the genes constitute different and independent datasets, application of this approach is over-conservative: as all alignments can have different lengths, sampling and divergence of taxa, the FDR ranking of the single tests according to p-values can be biased in favor of genes with more divergent taxa and/or higher amount of substitutions. This type of error is, however, conservative, as it can lead to a highly inflated rate of type 2 (false negative) errors, but keeps to minimum the rate of false negative errors.

In order to understand the influence of RNA editing in protein coding genes of molecular evolution, RNA editing sites were predicted in each gene alignments by PREPACT2 website (http://www.prepact.de/prepact-main.php). After that these sites were deleted in 45 genes for which RNA editing has been experimentally ascertained in some angiosperms, such as rubber tree, tobacco, pea and rice (Tangphatsornruang et al., 2011; Sasaki et al., 2003; Inada et al., 2004). At last, 11 families for whole gene, site and branch-specific predictions of positive selection for the revised alignments deleted RNA editing sites were tested with the same parameters.

3 Results

3.1 Phylogenetic reconstruction of the European *Aquilegia* rapid radiation by next-generation sequencing

3.1.1 Chloroplast genome assembly of *Aquilegia*

To expand our understanding of the relationships among Eurasian *Aquilegia* species, 34 individuals were sequenced through Illumina sequencing on Hiseq2000 instrument, producing 100bp paired end reads. According to the summary of Velvet *de novo* assembly for the 34 species showed in Table 3.1, several differences among these species were observed. First, the total number of reads after low-quality filtering were significantly different, e.g. *A. bertolonii* had the largest number of reads (3,102,847) but *A. ottonis* subsp. *speluncarum* had the least reads (266,249), and the average number of reads was 1,842,539. Second, most species had similar contig N50/N90 values, indicating that the quality of assembly among these species was stable, except for three species which had higher N50 values than others, *A. blecicii* (739/161), *A. ottonis* subsp. *speluncarum* (546/153) and *A. vulgaris* (520/137). Third, the total number of chloroplast contigs differed among taxa. Due to the lowest number of reads, *A. ottonis* subsp. *speluncarum* also had the lowest number of chloroplast contigs, but the total number of gene sequences were not influenced by the number of contigs, indicating that mainly non-coding sequences were affected by the uneven number of reads. Fourth, based on the gap-filled step of plastomes, *A. barbaricina* and *A. kitaibelii* had only three one-base gaps, and *A. nugorensis* had two one-base gaps. And the average length of plastomes without Ns, which contained the coding sequence information for this study, was about 134,764. To validate the accuracy of the assembled plastomes, the MAUVE alignments were carried out to compare plastomes with reference, indicating that the majority of sequences in whole plastomes were similar to the reference genome sequence except for some missing parts. The plastomes obtained were used for downstream analyses as detailed below.

Table 3.1: Assembly summary of 34 *Aquilegia* species

Species name	Number of reads	N50/N90 values	Number of chloroplast contigs	Total length of plastomes	Number of N base in plastomes
A. alpina	3025729	235/129	84	146875	5826
A. atrata	2970895	239/127	96	150907	31555
A. aurea	2821149	242/129	86	149810	22940
A. barbaricina	2797474	239/129	75	155373	3
A. bernardii	615254	302/161	49	148805	143
A. bertolonii	3102847	238/123	100	155137	17673
A. blecicii	1202717	739/161	74	147705	14519
A. dinarica	1561234	282/127	85	152834	447
A. dumeticola	564774	424/199	73	146750	15911
A. einseleana	2305147	256/126	77	152804	23646
A. grata	2597780	244/131	80	152922	37341
A. iulia	1809927	258/127	58	151852	14524
A. kitaibelii	572556	269/121	44	148149	3
A. litardierei	1032886	257/153	63	152390	33995
A. magellensis	1613778	256/138	58	152995	32058
A. nigricans	1924698	251/125	63	122639	3810
A. nugorensis	2533685	234/124	55	157442	2
A. nuragica	2559459	246/127	75	153112	9723
A. ottonis subsp. *amaliae*	2067852	253/124	82	151369	5755
A. ottonis subsp. *speluncarum*	266249	546/153	42	151560	432
A. pancicii	724762	302/155	56	152963	30619
A. pyrenaica	907682	349/168	61	152804	18184
A. reuterii	1963084	262/130	92	144453	4759
A. thalictrifolia	1075865	303/139	64	151823	2899
A. transsilvanica	1529952	267/140	91	150507	5960
A. vestinae	2485224	249/125	76	152099	29105
A. viscosa	2381238	285/136	118	148755	28393
A. vulgaris	1759584	520/137	85	154254	1907
A. buergeriana	2358922	227/134	84	151173	45777
A. glandulosa	2864165	234/142	171	146846	20584
A. olympica	1749194	285/139	84	127813	7619
A. oxysepala var. *oxysepala*	1813724	239/146	84	141022	920
A. sibirica	2421335	230/134	124	152996	9525
A. virdiflora	665499	318/185	63	146861	17250

3.1.2 Features of 66 CDS in *Aquilegia* genus

About 66 genes were chosen for molecular evolutionary analysis (Table 3.2), as they were those most commonly found among the taxa selected for analysis. Then they were aligned by the MACSE software and concatenated to create the final matrix for phylogenetic analyses. Among these genes, 4 genes (*ndhB, rpl2, rpl23, rps7*) were duplicated in inverted repeat regions (IRa and IRb). Other 8 genes were located in the SSC region while the LSC region contained 54 genes. A total of 9 genes included introns, and all of them were single intron genes except the *ycf3* gene, which was similar to *A. thaliana, Oryza sativa Japonica* and *C. resedifolia* (Hu et al., 2015). In addition, the *atpB* and *atpE* genes had a 3-bp overlapping region to each other, and the *psbC* and *psbD* genes had a 53-bp overlapping region.

Table 3.2: List of 66 CDS in the *Aquilegia* genus by DOGMA

Gene Category	Genes
Photosystem I	*psaA, psaB, psaC, psaI, psaJ*
Photosystem II	*psbA, psbC, psbD, psbE, psbF, psbH, psbI, psbJ, psbK, psbL, psbN, psbT, psbZ*
Cytochrome	*petA, *petB, *petD, petG, petL*
ATP synthase	*atpA, atpB, atpE, *atpF, atpH, atpI*
Rubisco	*rbcL*
NADH dehydrogenase	*ndhA,§*ndhB, ndhC, ndhD, ndhE, ndhF, ndhG, ndhH, ndhI, ndhJ, ndhK*
Ribosomal protein (large subunit)	§*rpl2, rpl14, *rpl16, rpl20, §rpl23, rpl33, rpl36*
Ribosomal protein (small subunit)	*rps3, rps4, §rps7, rps8, rps11, rps14, rps15, rps16, rps18, rps19*
RNA polymerase	*rpoA, rpoB, *rpoC1*
Cytochrome c biogenesis	*ccsA*
Membrane protein	*cemA*
Maturase	*matK*
Conserved reading frames	*ycf3, ycf4*

§Gene completely duplicated in the inverted repeat. *Gene with intron(s).

3.1.3 The GC content of chloroplast genes in the *Aquilegia* genus

The overall GC content of the 66 CDS were approximately 40%. Actually, the GC content of the first position in codon was about 48%, while at the second position around 40% and at the third position 32%. Then the GC content of genes was calculated by using EMBOSS website (http://www.bioinformatics.nl/emboss-explorer). Notably, the GC content of photosynthetic genes was higher than the one of ribosomal proteins and NADH dehydrogenase genes (Table 3.3). Photosynthetic genes contain rubisco large subunit, as well as photosystem I /II, cytochrome and ATP synthase subunit genes

(Shimda and Sugiuro, 1991). For instance, the average of the GC content of photosystem I genes was about 42.02%, whereas that of NADH dehydrogenase genes was 37.47%.

Table 3.3: GC content and that of the third position in codon of 66 CDS in *Aquilegia*

Genes	66 CDS	3rd	Genes	66 CDS	3rd
			Gene expression		
30S ribosomal proteins					
rps3	36.92	27.80	*rps14*	41.91	30.69
rps4	40.00	29.53	*rps15*	33.71	28.41
rps7	41.00	24.30	*rps16*	38.46	26.15
rps8	39.01	26.67	*rps18*	34.29	22.12
rps11	47.48	30.94	*rps19*	39.07	33.33
50S ribosomal proteins					
rpl2	45.82	36.73	*rpl23*	38.30	31.91
rpl14	39.56	26.81	*rpl33*	36.23	34.78
rpl16	45.52	31.33	*rpl36*	42.11	36.84
rpl20	39.27	30.51			
RNA polymerases					
rpoA	37.32	30.53	*rpoC1*	39.78	30.23
rpoB	41.42	33.78			
			NADH(P) dehydrogenase		
ndhA	36.36	25.01	*ndhG*	37.30	46.07
ndhB	37.70	31.31	*ndhH*	40.19	32.76
ndhC	36.63	28.93	*ndhI*	37.04	28.33
ndhD	36.18	29.95	*ndhJ*	42.75	38.31
ndhE	33.66	26.47	*ndhK*	40.79	32.46
ndhF	33.52	24.56			
			Photosynthesis		
Ribulose 1,5-diphosphate carboxylase/oxygenase					
rbcL	45.10	33.19			
Photosystem I					
psaA	43.72	35.42	*psaI*	37.84	32.43
psaB	42.50	35.42	*psaJ*	42.96	48.89
psaC	43.09	30.49			
Photosystem II					
psbA	43.77	38.10	*psbJ*	41.46	26.83
psbC	45.78	36.93	*psbK*	34.41	30.65
psbD	44.07	35.59	*psbL*	32.48	33.33
psbE	43.65	38.10	*psbN*	46.21	43.18
psbF	41.67	35.00	*psbT*	37.20	35.12
psbH	40.98	35.14	*psbZ*	33.33	22.22
psbI	37.74	34.02			
Cytochrome b/f comple					
petA	39.63	31.27	*petG*	39.47	36.84
petB	42.90	37.96	*petL*	35.42	25.00
petD	40.34	32.94			
cp H^+-ATPase					
atpA	42.32	32.09	*atpF*	38.20	34.05
atpB	43.55	32.65	*atpH*	46.75	28.05
atpE	40.30	32.09	*atpI*	38.85	29.03
			Other Genes		
cemA	34.50	35.68	*ycf3*	39.45	30.77
matK	33.01	26.57	*ycf4*	39.99	32.96
ccsA	33.13	28.31			

3.1.4 Amino acid polymorphisms of chloroplast genes in the *Aquilegia* genus

Besides the GC content analysis among 66 CDS, further analysis indicated that 20 genes had amino acid polymorphisms across taxa (Table 3.4). From this table, all the photosynthetic genes had one amino acid polymorphism except the *psbT* gene (two polymorphisms), indicating that the photosynthetic genes with high GC content are more conserved than other genes with low GC content. Six other genes (*ccsA, matK, ndhF, rpoB, rpoC1* and *rps7*) with low GC content had more than three amino acid polymorphisms, and the majority of polymorphisms among these genes occurred in first/third position of the codon. As to the NADH dehydrogenase genes, the *ndhF* gene with the lowest GC content and also lowest GC preferences in the third position had 11 polymorphisms, of which 5 polymorphisms were observed in the third site of codon position and 5 in the first site.

Table 3.4: Distribution of amino acid polymorphisms in 66 CDS of *Aquilegia*

Gene name	Codon position	Poly-morphisms	Site change	Gene name	Codon position	Poly-morphisms	Site change
atpB	120	Asn, His	1	*rpoB*	241	Glu, Asp	3
atpI	119	Ile, Val	1		569	Arg, Cys	1
psaB	247	Ser, Pro	1		588	Leu, Ile	1
psbA	123	Val, Ile	1	*rpoC1*	249	Asn, His	1
psbH	45	Val, Phe	1		571	Gln, Lys	1
psbI	52	Glu, Asp	3		573	Glu, Ala	2
psbT	1	Met, Ile	3		654	Ile, Lys	2
	2	Glu, Val	2	*rps7*	43	Leu, Ile	1
ndhA	106	Val, Ile	1		47	Leu, Met	1
ndhC	54	Ala, Thr	1		53	Glu, Lys	1
ndhD	22	Phe, Leu	3		68	Arg, Gly	1
ndhG	14	Phe, Leu	3		81	Ser, Gly	1
ndhH	185	Gly, Glu	2		106	Gly, Ala	2
ndhF	217	Asp, Tyr	1	*matK*	3	Glu, Lys	1
	365	His, Tyr	1		34	Ala, Ser	1
	509	Pro, Tyr	1		94	Leu, Phe	3
	653	Asn, Tyr	1		198	Thr, Ser	2
	733	Leu, Ile	1		204	His, Tyr	1
	525	Lys, Asn	3		264	Asn, His	1
	599	Met, Ile	3		370	Thr, Ile	2
	630	Leu, Phe	3		485	Ser, Tyr	2
	680	Leu, Phe	3	*ccsA*	169	Ile, Val	1
	709	Met, Ile	3		170	Asp, Glu	3
	475	Ser, Phe	2		207	Leu, Trp	2
rps4	199	Tyr, Asn	2		241	Val, Met	1
cemA	164	Pro,Thr	1		287	Gln, His	3

3.1.5 Codon usage analyses in *Aquilegia* and other five angiosperms

The codon usage frequency for all the *Aquilegia* species was calculated, and the average frequencies of 66 CDS in the *Aquilegia* genus were obtained. To compare the codon usage bias of *Aquilegia* to that of other five angiosperms, *Megaleranthis saniculifolia*, *Ranunculus macranthus*, *Thalictrum coreanum*, *Arabidopsis thaliana*, *Oryza sativa Japonica* (Figure 3.1), the same 66 chloroplast genes listed in Table 3.2 were calculated in these references for codon usage bias analysis. In this dataset, there were three stop codons (TAA, TAG and TGA), and all of amino acids had more than one codon except Tryptophan (TGG) and Methionine (ATG). The same amino acid with A or T in the third codon position was used more frequently than those terminating in G or C. Furthermore, all the amino acids had the same preferred codon usage in *Aquilegia* and the five reference species except Valine. In *Oryza sativa Japonica,* GTT (Val) codon had higher frequency than GTA (Val) codon, but the other five taxa had lower frequency of GTT codon than that of GTA codon. Among the five angiosperm species analyzed in this study, only *Oryza sativa Japonica* is a monocotyledon while the others are dicotyledons (Garnock-Jones, 1981; Ooka et al., 2003). One can speculate that Valine may play a role in codon usage bias for molecular evolution between eudicots and monocots.

Among these protein coding genes of 34 *Aquilegia* taxa, the majority of genes had ATG as start codon. Only the *ndhD* and *rpl2* genes had ACG as start codon, while in the *rps19* gene it was GTG. It is known that ACG, GTG or ATA can be used as an alternative to ATG as the start codon in basal eudicots, such as *Ranunculus*, *Nandina* and *Tetracentron* (Raubeson et al., 2007; Wu et al., 2014). It is generally known that RNA editing is essential for the normal functioning of the plant's translation and respiration activity. RNA editing including C-U editing and A-I editing occurs in the cell nucleus and cytosol, as well as within mitochondria and plastids (Danecek et al., 2012; Takenaka et al., 2014; Shikanai, 2015). Previous studies have illustrated that C-U editing lead to a specific translational start of *ndhD* in monocots, and translational efficiency of GUG codons is higher than AUG as start codon (Neckerman et al., 1994; Rohde et al., 1994) Thus further tests are needed to detect whether the alternative start codons in these genes will show a correlation with molecular evolution in *Aquilegia* clade or not.

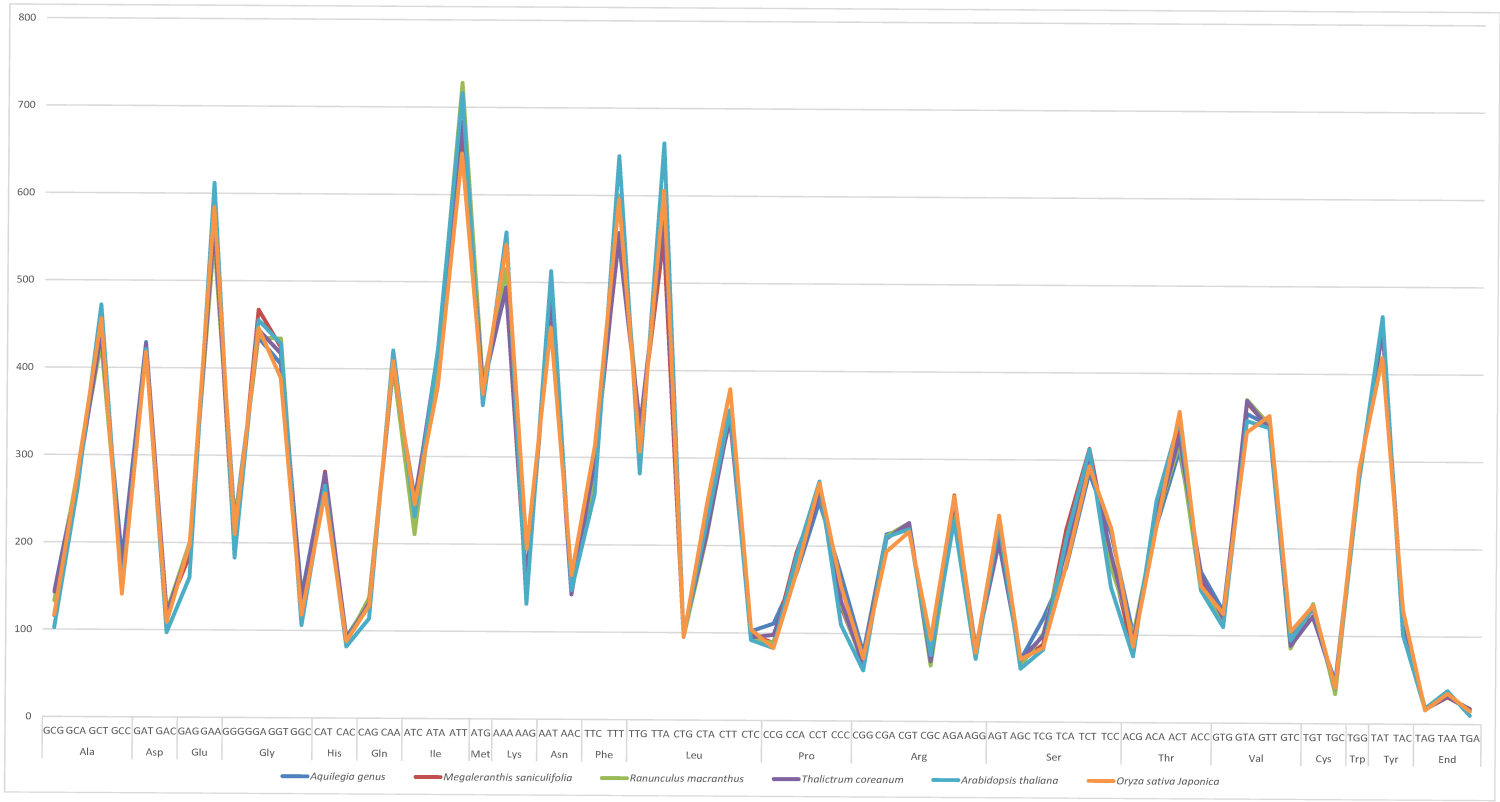

Figure 3.1: Codon usage bias in *Aquilegia* and other five angiosperms. The codon usage was calculated based on the same 66 chloroplast CDS among the *Aquilegia* genus and *Megaleranthis saniculifolia*, *Ranunculus macranthus*, *Thalictrum coreanum*, *Arabidopsis thaliana*, *Oryza sativa Japonica*.Y-axis reports the number of codon counts in 66 CDS for six species listed in X-axis.

3.1.6 Phylogenetic analyses of the European *Aquilegia* taxa

In this part of the study, four methods were applied to construct phylogenetic trees through 66 CDS dataset representing a total of 46194 nucleotides (Figure 3.2), and all the trees were re-rooted by *A. olympica* which had been consistently shown as basal to all other Eurasian *Aquilegia* taxa (Fior et al., 2013). Before constructing trees, the TPM1uf model was used as the best fitting model for the matrix by the test from the jModelTest program. CodonPhyML indicated GY as codon model to obtain a fast maximum likelihood of phylogenetic inference, while PhyML applied a custom (012345) model to estimate maximum likelihood phylogenies from alignments of nucleotide sequences with a -lnL of 64188.5172. The ML trees inferred by CodonPhyML and PhyML had very similar topologies (Figure 3.2A, B), except for some localized differences, such as the branches of *A. transsilvanica* and *A. aurea*, and the branches of *A. bernardii* and *A. litardierei*. The MP tree reconstructed by PHYLIP, only separated Asian and European clades (Figure 3.2C). The highest support values and the best distinctness of phylogenetic relationships were gained through bayesian inference using the Mrbayes program (Figure 3.2D).

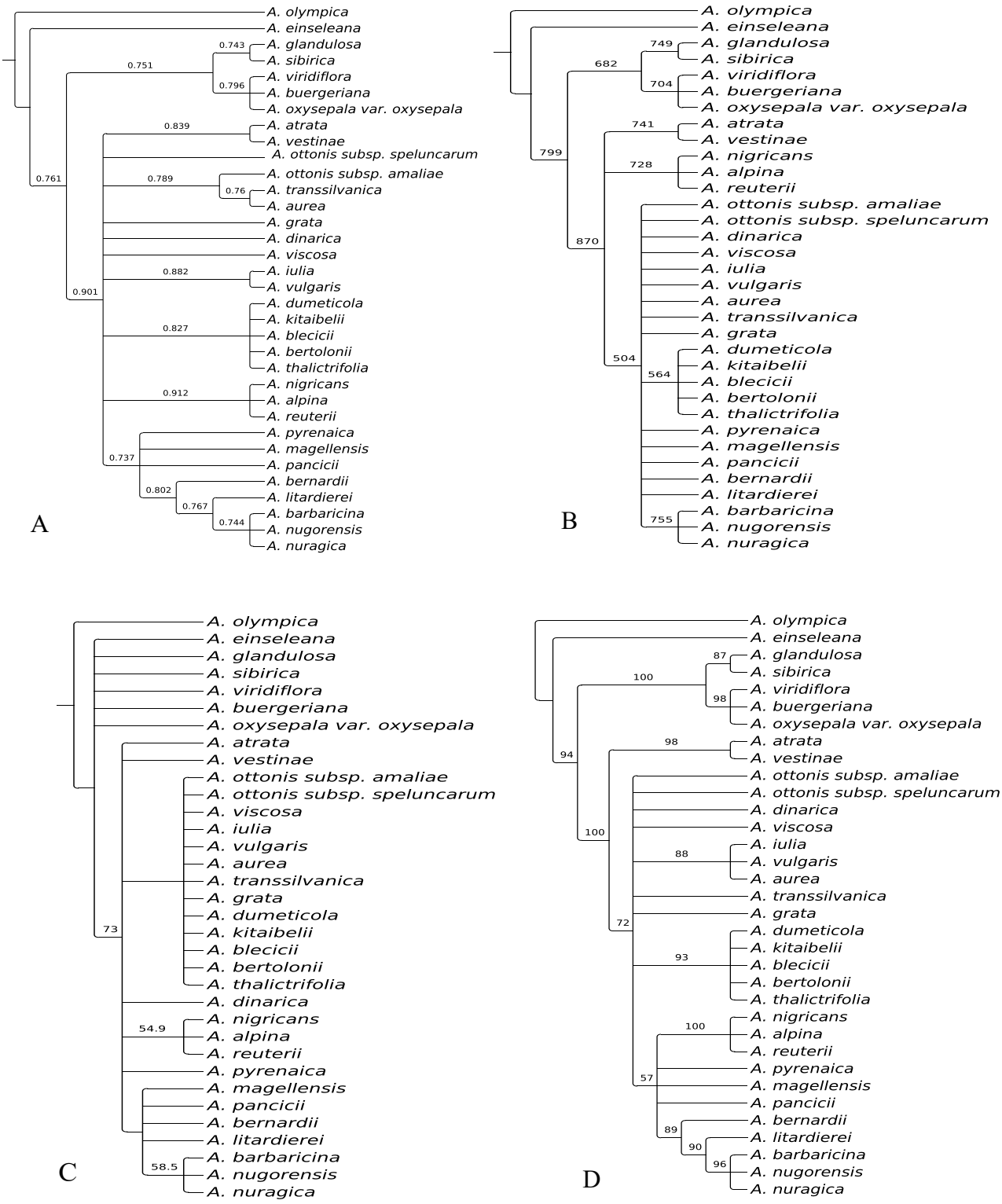

Figure 3.2: Phylogenetic trees among the *Aquilegia* taxa for CDS sequences dataset by four methods. A represents tree constructed by CodonphyML; B demonstrates PhyML tree; C illustrates PHYLIP tree; D indicates MrBayes tree.

Based on the topology of the Mrbayes, a single phylogenetic tree was constructed to display the results combined by Mrbayes, PHYLIP and PhyML methods through TreeGraph2 program (Figure 3.3). The clade encompassing the taxa from *A. olympica* to *A. oxysepala* var. *oxysepala* were Asian, except *A. einseleana,* and this relationships of subtree was supported by Fior's publication (Fior et al., 2013). All the Asian species were separated from European species with high support values through this dataset, even though *A. viridiflora, A. buergeriana* and *A. oxysepala* var. *oxysepala* were poorly resolved from each other in the Asian clade. Besides, the other European taxa were closely related, but this figure elaborated incomplete relationships between these species in this phylogeny, because the branches among several species were not separate, such as the branches between *A. ottonis* subsp. *amaliae* and *A. grata*, *A. dumeticola* and *A. thalictrifolia*. With regard to the low support value from *A. nigricans* to *A. nuragica,* this branch was different between two methods (Mrbayes and PhyML). The main reason causing some lack of resolution must be likely attributed to the fact that the taxa were highly conserved, and did not reveal obvious differences in evolutionary positions in the phylogenetic tree. Previous studies demonstrated that the morphology of leaves, stems and roots of all taxonomic units were very similar to each other (Prażmo, 1965), making it impossible to clarify relationships based on this information. Therefore, these results highlight that it is possible to obtain a large amount of phylogenetic information from chloroplast sequences, which in turns allows to improve the accuracy of phylogenetic relationships definition among *Aquilegia* species.

To precisely determine the phylogenetic relationships among Eurasian *Aquilegia*, whole plastome sequences (incomplete, 98314bp) were utilized as another dataset to analyze the evolutionary relationships of 34 *Aquilegia* species, in the hope that the additional information contained in non-coding regions may help to further resolve European taxa. The same three methods (Mrbayes, PHYLIP and PhyML) were applied to reconstruct phylogenetic trees for improving the results of the CDS dataset. This new dataset involved both protein coding sequences and non-coding sequences which included intronic regions and intergenic regions. A previous study showed that coding sequences combined with non-coding sequences significantly increased support value and resolution for phylogenies (Bremer et al., 2002). Therefore, the higher accuracy of phylogenetic tree from the whole plastome dataset can most likely be attributed to its larger amount of sequence information. The summary tree of these analyses reported in Figure 3.4 shows that, owing to the increased amount of information from non-coding sequences, more detailed results were obtained. For instance, the subtree of Asian taxa which included *A. glandulosa*, *A. sibirica*, *A. viridiflora, A. buergeriana* and *A. oxysepala* var. *oxysepala* had very high support values by all the three methods. But another Asian taxa, *A. olympica,* was separated from them because of *A. einseleana* insertion, consistently to what previously observed (Fior et al., 2013). With regard to *A. einseleana*,

a European species, it divided Asian taxa into two parts with low support values, but the evolutionary position in MP and ML methods were significantly different from the Mrbayes method. Even if the final evolutionary position of *A. einseleana* was uncertain due to the low support values, this species was more closely related to Asian taxa than other European taxa.

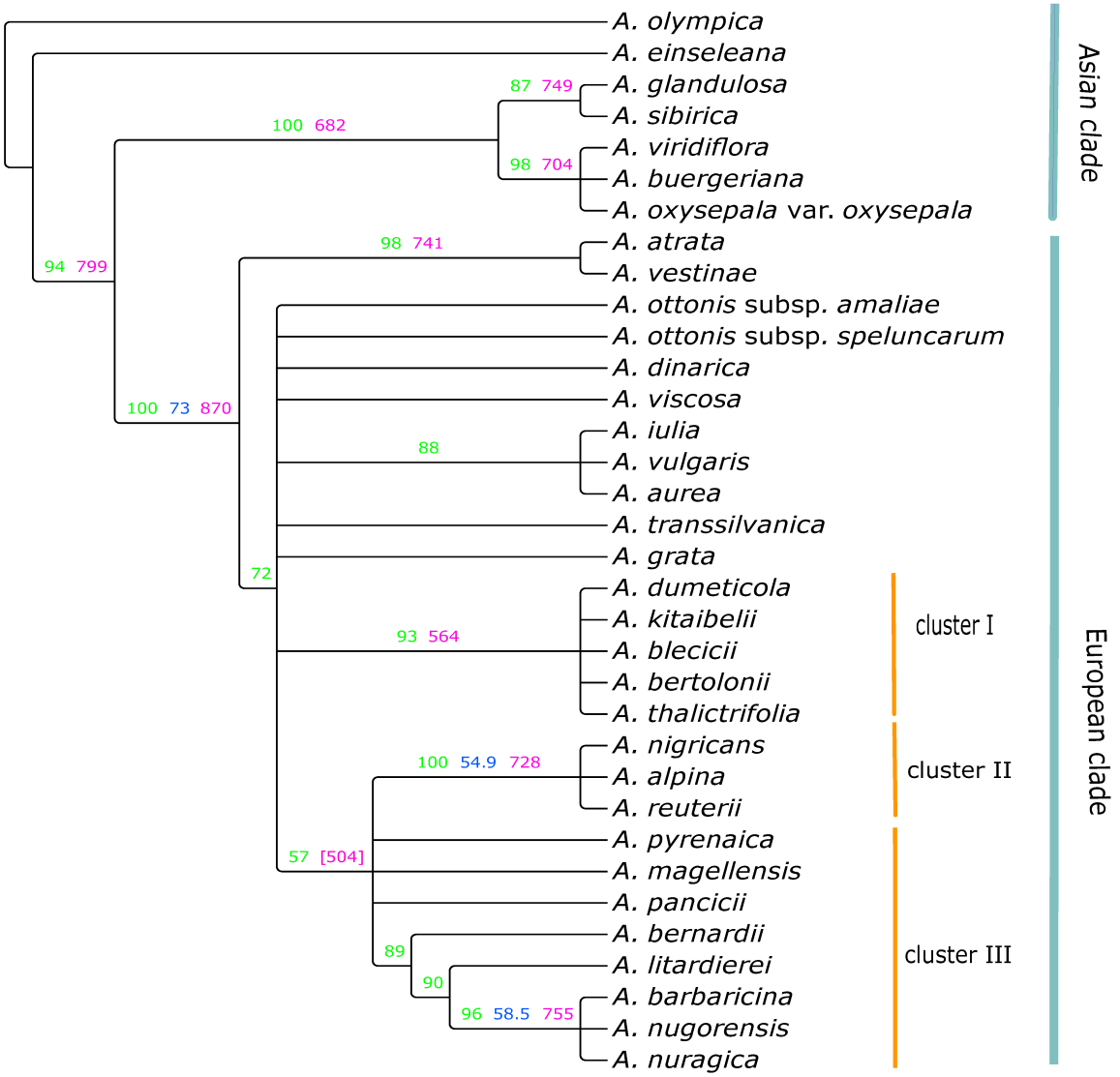

Figure 3.3: Cladogram of the phylogenetic relationships among *Aquilegia* for CDS sequences dataset. The cladogram represents the consensus topology of BI, MP and ML. Green color represents BI support values, blue color means MP support values, and purple color indicates ML support values.

Additionally, there were three important clusters which belonged to European taxa in this phylogenetic tree. The first cluster contained *A.dumeticola, A.kitaibelii, A.blecicii, A.thalictrifolia* and *A.bertolonii,* and it demonstrated that these species were closely related in evolution, as supported by both of the datasets. But the result of the whole plastome dataset had an obvious advantage compared to the CDS dataset. The second cluster included *A. nigricans*, *A. alpina* and *A. reuterii*, and they had extremely high support values. Compared with previous studies (Fior et al., 2013; Li et al., 2014), this result represented a significant improvement because of the evolutionary branches among these clades. The third cluster involved *A. pyrenaica, A. magellensis, A. pancicii, A. bernardii, A. litardierei, A. barbaricina, A. nugorensis* and *A. nuragica*. Although *A. pyrenaica* and *A. magellensis* were still merged to each other, more precise relationships between *A. pancicii* and *A. nuragica* were obtained in contrast to the CDS tree. With respect to the subtree constituted by *A. barbaricina, A. nugorensis* and *A. nuragica*, all of them were collected from Sardegna, the second largest Italian island. Therefore, these three species were closely related to each other, thus explaining the lack of resolution to study the phylogenetic relationships among them based only on chloroplast genome data.

Furthermore, the other two subtrees (*A. atrata* and *A.vestinae*, *A. ottonis* subsp. *amaliae* and *A. ottonis* subsp. *speluncarum)* had the same evolutionary position as occurred in the previous studies (Fior et al., 2013; Li et al., 2014). In fact, the subspecies of *A. ottonis* (Guacchio, 2009) had a very close relationship in the phylogeny, thus this result was correct even though they were supported by low values apart from other species. Besides, as to the other seven species (*A. dinarica, A. viscosa, A. iulia, A. vulgaris, A. aurea, A. transsilvanica* and *A. grata*), more potential sequence information of chloroplast DNA sequences or mitochondrial/nuclear sequence will be needed to improve the phylogenetic reconstruction.

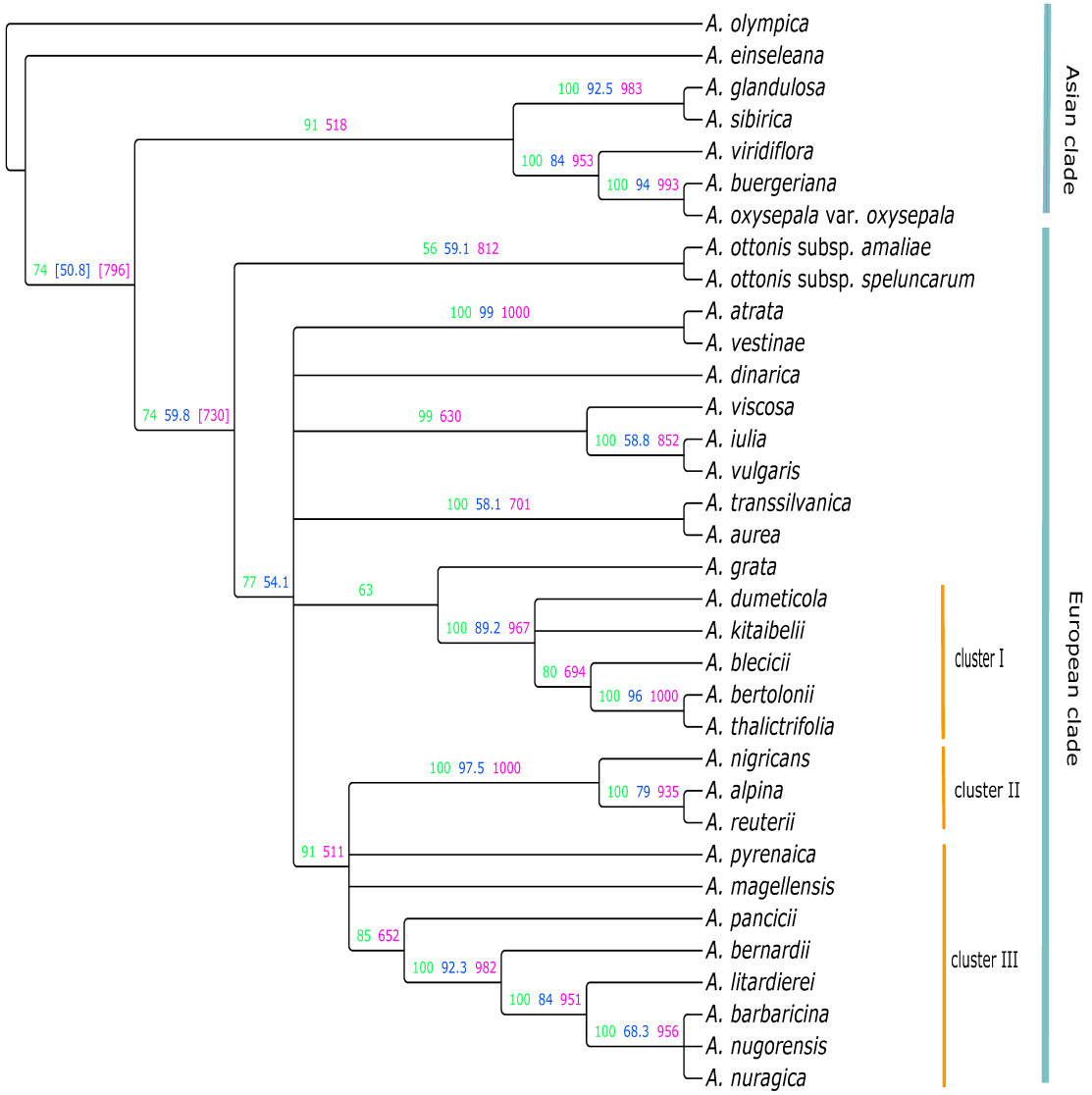

Figure 3.4: Cladogram of the phylogenetic relationships among *Aquilegia* for whole plastomes dataset. The cladogram represents the consensus topology of BI, MP and ML. Green color represents BI support values, blue color means MP support values, and purple color indicates ML support values.

3.2 Reconstructing the tribal relationships in the Brassicaceae family

3.2.1 Chloroplast genome assembly of Brassicaceae

To expand our understanding of the relationships among the Brassicaceae family, 78 individuals were sequenced through Illumina sequencing on Hiseq2000 instrument, producing 100bp paired end reads. According to the summary of Velvet *de novo* assembly for the 78 species showed in Table 3.5, the total number of reads after filtering low-quality step was significantly different, e.g. *Cochlearia officinalis* had the largest number of reads (3,915,604) while *Diplotaxis tenuifolia* had the least reads (1,001,488), and the average number of reads across species was 2,504,441. The number of contigs for all the species had a range from 327 to 20,649, but their mean was 3,834.75. Additionally, the size of predicted whole plastomes after gap-filling were obviously different, from 111,816 base pairs to 162,131 base pairs. Besides, all of them had high depth of coverage per plastome. To validate the accuracy of the assembled plastomes, MAUVE alignments were carried out to compare plastomes with the reference, indicating that the majority of sequences in whole plastomes were similar to reference except for some missing parts. In this study, 71 CDS from 95 Brassicaceae species were extracted for the following study in phylogenetic and evolutionary analyses.

Table 3.5: Summary of assembly for 78 species in Brassicaceae

Species name	Number of reads	Number of contigs	Depth of coverage	Total length of plastomes
Berteroa incana	2999844	1942	977.37	128457
Alyssum alissoides	3062700	5153	997.85	144812
Fibigia clypeata	2728212	3863	888.87	156087
Matthiola fruticulosa	2594036	3028	845.15	128868
Bunias orientalis	2629820	4544	856.81	128464
Draba verna	1941988	1857	632.71	153914
Draba dubia	1887084	6190	614.82	137886
Arabis alpina	2438728	3149	794.55	153890
Arabis hirsuta Aggreg	1569304	5327	511.29	156459
Arabis nova	3042836	1238	991.38	153529
Arabis soyeri subsp. *subcoriacea*	1418444	4523	462.14	128243
Arabis turrita	3087084	2434	1005.79	131848
Boechera gracilipes	3349244	2466	1091.21	155429
Phoenicaulis cheiranthoides	2593912	3663	845.11	129552
Polyctenium fremontii	2548096	3531	830.19	154981
Diplotaxis tenuifolia	1001488	5119	326.29	133661
Brassica repanda subsp. *baldensis*	2404168	2947	783.29	133577
Hirschfeldia incana	1145660	3014	373.26	146737
Camelina microcarpa	2394012	1974	779.98	155554
Capsella grandiflora	1811176	865	590.09	155306
Erysimum aurantiacum	1480972	4124	482.51	149916
Erysimum rhaeticum	2518844	20107	820.66	154424
Erysimum sylvestre	2817132	4374	917.84	154459
Erysimum virgatum	2102120	3718	684.88	138547
Neslia paniculata	2340736	1729	762.63	129431
Rorippa sylvestris	2497488	3362	813.70	154437
Cardamine hirsuta	3099920	2877	1009.97	159840
Cardamine alpina	2500352	2265	814.63	130179
Cardamine flexuosa	2914608	2751	949.60	133018
Rorippa austriaca	1754756	2213	571.71	129067
Dentaria enneaphyllos	3047592	5505	992.93	130418
Dentaria pentaphyllos	2074116	8077	675.76	150642
Leavenworthia uniflora	2763736	1069	900.44	155675
Leavenworthia exigua	3344420	3242	1089.63	129360
Cochlearia officinalis	3915604	5005	1275.73	137123
Descurainia bourgaeana	1922576	1455	626.39	147834
Descurainia sofia	2790740	327	909.24	128917
Hornungia petraea	2150672	1151	700.70	139713
Hutchinsia alpina	2896852	20649	943.81	161800
Hutchinsia brevicaulis	2633200	2262	857.91	155251

Hymenolobus pauciflorus	2103652	2427	685.38	154327
Malcolmia littorea	3207388	4098	1044.99	131467
Morettia philaeana	1695472	507	552.40	128582
Thellungiella halophila	2828928	1662	921.68	154310
Halimolobos pubens	3475880	3959	1132.46	130113
Heliophila coronopifolia	3117228	1098	1015.61	154245
Hesperis matronalis	2980812	4698	971.17	154767
Iberis amara	1297980	3558	422.89	147147
Isatis tinctoria	2878588	2185	937.86	128148
Lepidium campestris	2935484	2705	956.40	129598
Cardaria draba	1062324	809	346.11	140835
Noccaea precox	3387912	3723	1103.80	128861
Noccaea rotundifolium	3114292	6053	1014.66	127776
Lesquerella montana	2056408	4967	669.99	154409
Nerisyrenia camporum	1962748	1487	639.48	154838
Stanleya pinnata	2861628	1659	932.34	129562
Thelypodium laciniatum	2720716	3819	886.43	153465
Ochthodium aegyptiacum	1695540	5429	552.42	128625
Sisymbrium officinale	2414628	4781	786.70	128705
Smelowskia calycina	3414796	2689	1112.56	111816
Thlaspi perfoliatum	2874836	356	936.64	127979
Peltaria angustifolia	1610204	3516	524.61	129661
Biscutella laevigata	3122740	4543	1017.41	155726
Biscutella prealpina	1625844	4863	529.71	136671
Calepina irregularis	1658292	1372	540.28	155009
Kernera saxatilis	1506884	10903	490.95	128574
Lunaria annua	1173784	3222	382.43	160455
Cleome spynosa	3045992	4758	992.40	158130
Cleome hirta	1116868	2360	363.88	162131
Alyssum dasycarpum	3088728	4210	1006.33	127468
Draba aizoides	3284752	2786	1070.19	127707
Turritis glabra	2953208	4352	962.17	154562
Cardamine pentaphyllos	3289460	4913	1071.73	133255
Cardamine asarifolia	2551672	6145	831.35	154399
Cardamine trifolia	3127504	3429	1018.96	150884
Cardamine pratensis	3416204	6146	1113.02	126670
Aethionema saxatile	2977624	2595	970.13	157400
Arabidopsis halleri	3495144	3240	1138.74	155607

3.2.2 Features of 71 CDS in the Brassicaceae family

After annotating, 71 CDS from 78 sampled species and 17 reference species were chosen for molecular evolutionary analysis (Table 3.6), as they were those most commonly found in common among the taxa selected for analysis. These genes were aligned by the MACSE software and concatenated to create the final matrix for phylogenetic analyses. Among these genes, 5 genes (*ndhB, rpl2, rpl23, rps7, rps12*) were duplicated in inverted repeat regions (IRa and IRb). Only 10 genes were located in the SSC region while the LSC region included 56 genes. A total of 10 genes had introns, and all of them were single intron genes except the *ycf3* gene. Furthermore, these genes were divided into several categories according to functions, such as Photosystem I/II, Cytochrome, ATP synthase, NADH dehydrogenase, Ribosomal protein.

Table 3.6: List of 71 CDS in the Brassicaceae family

Gene Category	Genes
Photosystem I	*psaA, psaB, psaC, psaI, psaJ*
Photosystem II	*psbA, psbB, psbC, psbD, psbE, psbF, psbH, psbJ, psbL, psbM, psbN, psbT, psbZ*
Cytochrome	*petA, *petB, *petD, petG, petL, petN*
ATP synthase	*atpA, atpB, atpE, *atpF, atpH, atpI*
Rubisco	*rbcL*
NADH dehydrogenase	*ndhA,§*ndhB, ndhC, ndhD, ndhE, ndhF, ndhG, ndhH, ndhI, ndhJ, ndhK*
Ribosomal protein (large subunit)	§*rpl2, rpl14, *rpl16, rpl20, rpl22, §rpl23, rpl33, rpl36*
Ribosomal protein (small subunit)	*rps2, rps3, rps4, §rps7, rps8, rps11, §*rps12, rps14, rps15, rps18, rps19*
RNA polymerase	*rpoA, rpoB, *rpoC1, rpoC2*
ATP-dependent protease	**clpP*
Cytochrome c biogenesis	*ccsA*
Membrane protein	*cemA*
Maturase	*matK*
Conserved reading frames	*ycf4*
Pseudogenes	*accD*

§Gene completely duplicated in the inverted repeat. *Gene with intron(s).

3.2.3 Positive selection analysis among protein coding genes

Detecting biologically significant amino-acid sites is important, e.g. for the study of drug design or protein function. Conserved sites are indicative of functionally active sites (Drory et al., 2004) or protein-protein interaction epitopes while highly variable sites may represent sites subjected to positive Darwinian selection (Zhang et al., 2005). Such positively selected sites may be interpreted as being a consequence of molecular

adaptation, which confers an evolutionary advantage to the organism (Busset et al., 2011). In this study, Selecton was ran for both the null and positive selection model for the 71 CDS. This resulted in the prediction of 33 genes under positive selection (degree=1, score of probability ≤ 0.05). Based on functional gene categories, all the categories had genes detected as under positive selection except photosystem I (Figure 3.5). And the largest number of these positively selected genes were concentrated in "Transcription and Translation" (ribosomal protein subunits and RNA polymerase gens) and "Electron transport and ATP synthesis" (NADH dehydrogenase and ATP synthase genes).

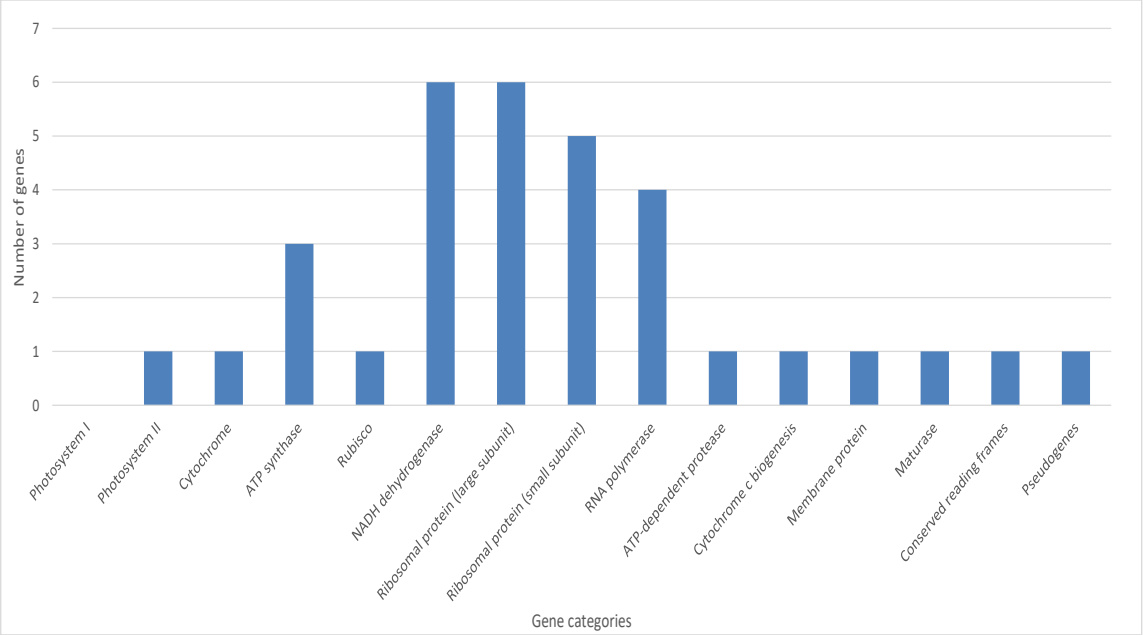

Figure 3.5: Gene number under positive selection in different categories by Selecton. Y-axis reportes the number of genes under positive selection in each of the functional classes listed on the X-axis.

3.2.4 Phylogenetic analyses of Brassicaceae

In this study, phylogenetic analyses including codonPhyML, PhyML, PAUP, RaxML and Mrbayes programs were carried out for the 71 CDS of 95 species in the Brassicaceae family. All the trees were re-rooted by the Cleomaceae family which was the sister to Brassicaceae in a previous study (Schranz and Mitchell-Olds, 2006). Before reconstructing the phylogenetic trees, the jModeltest and ProtTest programs were used to determine parameters of nucleotide substitution and amino acid substitution for all phylogenetic analyses, respectively. As to the nucleotide substitution, CodonPhyML analysis selected GY as substitution model and F3X4 as frequency model. PhyML analysis based on the GTR+I+G model resulted in a single tree with lnL= -309891.89. Bootstrap analysis indicated that 82 of 93 nodes were supported by values ≥ 95% and 80

of them were with bootstrap values of 100%. RAxML analysis utilized GTR model with 100 bootstraps as a fast method to launching a large phylogenetic trees based on inference. PAUP analysis resulted in a single tree, with a consistency index of 0.546, and a retention index of 0.678. Bootstrap analysis indicated that 82 of 90 nodes were supported by values ≥ 95% and 79 of them were with bootstrap values of 100%. Mrbayes analysis resulted in a single tree based on GTR+I+G model, and 87 of 93 nodes were supported by posterior probability values ≥ 0.95 and 84 of them were with posterior probability values of 1. The topologies of trees resulting from parsimony, likelihood and bayesian analyses were statistically not significantly different in this study.

Figure 3.6 displays the phylogenetic relationships of Brassicaceae based on the nucleotide substitution of 71 chloroplast CDS by combining the results of Mrbayes, Paup, RAxML, CodonPhyML and PhyML methods (Figure 3.7 based on amino acid substitution). These two figures demonstrated that the tribes of Brassicaceae were monophyletic and distinct from the outgroup taxa, but the results based on amino acid substitution had lower support values than those of nucleotide substitution. The outgroup, *Cleome spinosa* and *Cleome hirta* which belonged to the Cleomaceae family were chosen based on former literature as the root in this tree (Mithen et al., 2010). The basal tribe of Brassicaceae was Aethionemeae, including *Aethionema saxatile* and other two references (*Aethionema grandiflorum* and *Aethionema cordifolium*). Besides, three major lineages were defined in this phylogeny. Lineage I consisted of Camelineae, Turritideae, Erysimeae, Alyssopsideae, Microlepidieae, Crucihimalayeae, Boechereae, Halimolobeae, Physarieae, Cardamineae, Lepidieae, Descurainieae and Smelowskieae. It comprised the species from *Capsella bursa-pastoris* to *Smelowskia calycina* in this tree. Lineage II consisted of Brassiceae, Sisymbrieae, Thelypodieae and Isatideae, which involved the species from *Brassica napus* to *Isatis tinctoria*. Lineage III consisted of Biscutelleae, Anchonieae, Hesperideae and Buniadeae, that included species from *Biscutella laevigata* to *Bunias orientalis*. In fact, there had an expanded lineage II which included Calepineae, Eutremeae, Thalaspideae, Arabideae, Iberideae, Cochlearieae, Heliophileae, Kernereae, Alysseae, and the species were from *Calepina irregularis* to *Thlaspi perfoliatum*. The other two species, *Lunaria annua* and *Ochthodum aegyptiacum* were not assigned to any tribe.

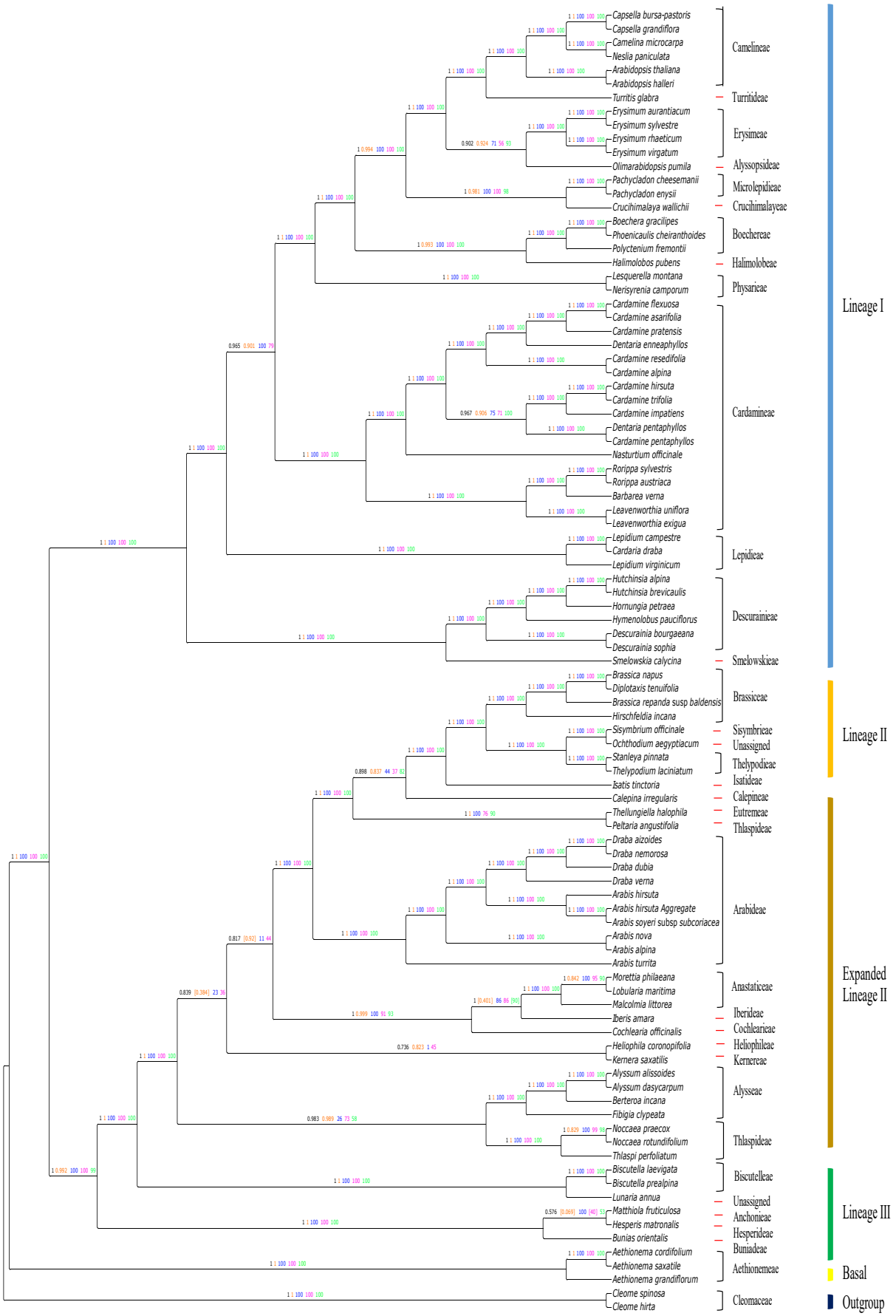
Capsella bursa-pastoris
Capsella grandiflora
Camelina microcarpa
Neslia paniculata
Arabidopsis thaliana
Arabidopsis halleri
Camelineae
Turritis glabra
Turritideae
Erysimum aurantiacum
Erysimum sylvestre
Erysimum rhaeticum
Erysimum virgatum
Erysimeae
Olimarabidopsis pumila
Alyssopsideae
Pachycladon cheesemanii
Pachycladon enysii
Microlepidieae
Crucihimalaya wallichii
Crucihimalayeae
Boechera gracilipes
Phoenicaulis cheiranthoides
Polyctenium fremontii
Boechereae
Halimolobos pubens
Halimolobeae
Lesquerella montana
Nerisyrenia camporum
Physarieae
Cardamine flexuosa
Cardamine asarifolia
Cardamine pratensis
Dentaria enneaphyllos
Cardamine resedifolia
Cardamine alpina
Cardamine hirsuta
Cardamine trifolia
Cardamine impatiens
Dentaria pentaphyllos
Cardamine pentaphyllos
Nasturtium officinale
Cardamineae
Rorippa sylvestris
Rorippa austriaca
Barbarea verna
Leavenworthia uniflora
Leavenworthia exigua
Lepidium campestre
Cardaria draba
Lepidium virginicum
Lepidieae
Hutchinsia alpina
Hutchinsia brevicaulis
Hornungia petraea
Hymenolobus pauciflorus
Descurainia bourgaeana
Descurainia sophia
Descurainieae
Smelowskia calycina
Smelowskieae
Lineage I
Brassica napus
Diplotaxis tenuifolia
Brassica repanda susp baldensis
Hirschfeldia incana
Brassiceae
Sisymbrium officinale
Sisymbrieae
Ochthodium aegyptiacum
Unassigned
Stanleya pinnata
Thelypodium laciniatum
Thelypodieae
Isatis tinctoria
Isatideae
Calepina irregularis
Calepineae
Thellungiella halophila
Eutremeae
Peltaria angustifolia
Thlaspideae
Lineage II
Draba aizoides
Draba nemorosa
Draba dubia
Draba verna
Arabis hirsuta
Arabis hirsuta Aggregate
Arabis soyeri subsp subcoriacea
Arabis nova
Arabis alpina
Arabis turrita
Arabideae
Morettia philaeana
Lobularia maritima
Malcolmia littorea
Anastaticeae
Iberis amara
Iberideae
Cochlearia officinalis
Cochlearieae
Heliophila coronopifolia
Heliophileae
Kernera saxatilis
Kernereae
Alyssum alissoides
Alyssum dasycarpum
Berteroa incana
Fibigia clypeata
Alysseae
Noccaea praecox
Noccaea rotundifolium
Thlaspi perfoliatum
Thlaspideae
Biscutella laevigata
Biscutella prealpina
Biscutelleae
Lunaria annua
Unassigned
Matthiola fruticulosa
Anchonieae
Hesperis matronalis
Hesperideae
Bunias orientalis
Buniadeae
Expanded Lineage II
Aethionema cordifolium
Aethionema saxatile
Aethionema grandiflorum
Aethionemeae
Cleome spinosa
Cleome hirta
Cleomaceae
Lineage III
Basal
Outgroup

Figure 3.6: Cladogram of the phylogenetic relationships among 95 Brassicaceae species of 71 CDS based on nucleotide substitution. The cladogram represents the consensus topology of BI, MP and ML. Numbers on branches indicate different support values. Black color illustrates Mrbayes support values, orange color shows codonphyml support value, blue color represents raxml support values, purple color means phyml support values and green color indicates paup support values.

The Cleomaceae is a small family of flowering plants in the order Brassicales, comprising about 150 species in 7 genera. The APG II system allowed for Cleomaceae to be included in the Brassicaceae family in 2003. In this study, Cleomaceae clade had strong support values (1/1/100/100/100) as the root to all other tribes of Brassicaceae, and this conclusion was also sustained by other studies (Beilstein et al., 2006; Beilstein et al., 2008; Franzke et al., 2009). The tribe of Aethionemeae which has 56 species and originates from sunny limestone mountainsides in Europe and West Asia was the sister with Cleomaceae in this tree, occupying the "basal" position with respect to the rest of the Brassicaceae species. Although only *Aethionema saxatile* was sequenced and other two reference plastomes were analyzed in the *Aethionema* genus, the topologies from different approaches provided high support for the same relationships between this tribe to the others.

Lineage I was a well-supported monophyletic group including 48 species from 13 tribes, and characterized by the presence of forked and dendritic trichomes. The tribe of Camelineae included 12-13 genera and approximately 240 species distributed primarily in Eurasia, but in this study it only had 4 genera and 6 species. *Capsella grandiflora* and *Capsella bursa-pastoris* formed a monophyletic group, *Capsella,* which was sister to *Camelina* and *Neslia* genera. *Turritis glabra*, Turritideae, was strongly supported as closely relative to Camelineae. Eryismeae had about 220 species, which was supported as the sister of *Olimarabidopsis pumila* (Alyssopsideae). The relationships among these four tribes were sustained in Beilstein study (Beilstein et al., 2006). The tribes of Microlepidieae and Crucihimalayeae in this study only included three reference species which had been observed to belong to lineage I in previous studies (Huang et al., 2015; Koch et al., 2007), and they were found to be close to each other in this phylogeny. There were two new tribes in this lineage, Boechereae and Halimolobeae. Boechereae includes 7 genera and about 110 species, most of which belong to *Boechera*, while Halimolobeae contains 5 genera and 40 species. Both of them were strongly supported as sisters to each other in this phylogeny. The Physarieae tribe was first identified as a monophyletic clade in 2003, which consists of 7 genera and 150 species. This tribe is readily distinguished from the other tribes of the Brassicaceae by having pollen with four or more colpi (all the others are tricolpate). As to the Cardamineae, it has over 340 species from 10 genera, and sequences from *Barbarea, Cardamine, Nasturtium, and Rorippa* are available for the

current studies (Beilstein et al., 2006; Beilstein et al., 2008; Huang et al., 2015). The Lepidieae consists of 3-5 genera and over 240 species, and the core genus is *Lepidium*. In addition, both of Cardamineae and Lepidieae are distributed on all continents except Antarctica (Al-Shehbaz et al., 2006; Franzke et al., 2009). Another new tribe, Descurainieae, has about 6 genera and 60 species distributed in the Americas, Eurasia, and Africa. And this tribe was strongly supported as the sister of Smelowskieae which is a new unigeneric tribe. In this phylogeny, the topology from Boechereae to Smelowskieae not only supported the relationship among tribes in previous studies, but also provided a deeper understanding in these tribes than previous researches (Al-Shehbaz et al., 2006; Franzke et al., 2009).

Lineage II in this study included Brassiceae, Sismbrieae, Thelypodieae, Isatideae, and even an unassigned species (*Ochthodium aegyptiacum*). It is generally known that most of studies has sustained that Schizopetaleae, Brassiceae, Sismbrieae and Isatideae are in lineage II for Brassicaceae classification (Beilstein et al., 2006; Al-Shehbaz et al., 2006; Beilstein et al., 2008; Franzke et al., 2009). Schizopetaleae is the earliest name, and the most common name is Thelypodieae. Thus, this phylogeny also placed these four tribes in lineage II. The most important tribe, Brassiceae, consists of 46 genera and about 230 species, and it is considered as an essential object in molecular studies because of the economically important Brassica species and its relatives. Sismbrieae consists of 70 genera and 400 species, but *Sisymbrium* only has about 40 species distributed in Eurasia and Africa. Thelypodieae consists of about 230 species in at least 20 genera, and it is well supported as the sister of Sismbrieae. Besides, Isatideae has over 90 species in 8 genera, and it forms a monophyletic group based strictly on morphology (Koch et al., 2003).

Lineage III included Biscutelleae, Anchonieae, Hesperideae, Buniadeae, and even an unassigned species (*Lunaria annua*). These four tribes had been defined in lineage III in previous studies (Beilstein et al., 2006; Beilstein et al., 2008; Franzke et al., 2009; Franzke et al., 2011; Huang et al., 2015). The Anchonieae tribe contains 12 genera and about 130 species distributed primarily in Eurasia and eastern and northern Africa, and only few species in North America. The unigeneric tribe, Hesperideae, consists of *Hesperis* (46 species) which is a genus centered in the Middle east and Europe. The tribe of Buniadeae has only one genus (*Bunias*) which contains only two accepted species, *Bunias erucago* and *Bunias orientalis*. In this phylogeny, the above three tribes were strongly supported to belong to lineage III and closely related to Aethionemeae tribe, but the relationships among them were not completely solved. The unassigned species, *Lunaria annua*, was well supported as the sister of Biscutelleae, and this situation had been observed in other studies (Beilstein et al., 2006; Huang et al., 2015).

Expanded lineage II was firstly defined by Franzke (Franzke et al., 2011). A group which contains a number of paraphyletic taxa near lineage II is named expanded lineage II. All the tribes defined belonging to expanded lineage II in this phylogeny were

supported by previous studies except the Anastaticeae tribe (Franzke et al., 2011; Huang et al., 2015). Eutremeae and Thlaspideae were closely related to each other in this study, and its relationship had been sustained in other studies (Beilstein et al., 2008; Franzke et al., 2009). Thlaspideae consists of 26 species in 7 genera, which is restricted to Europe and southwestern Asia. In addition, Alysseae was sister of Thlaspideae supported in Koch's research (Koch et al., 2007), where *Noccaea* was defined as *Thlaspi*. The majority of *Alyssum* and at least some of the genera formed a monophyletic group, Alysseae. The Arabideae comprises at least 6 genera and over 460 species distributed primarily in Eurasia and North America. Despite being the largest genus in this family and one of the most diversified morphologically, *Draba* (440 species) is a monophyletic genus in Arabideae (Warwick et al., 2006).

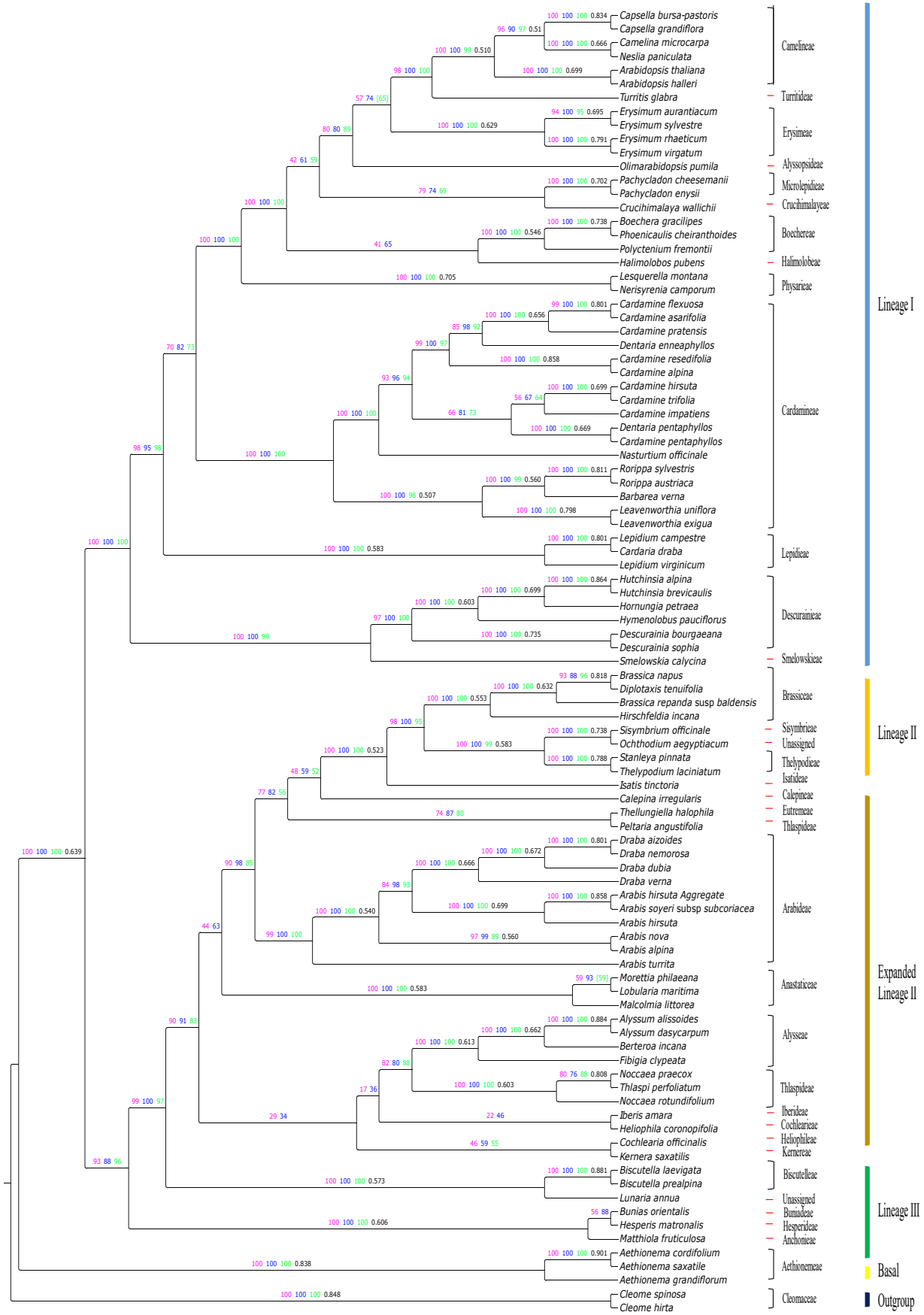
Capsella bursa-pastoris
Capsella grandiflora
Camelina microcarpa
Neslia paniculata
Arabidopsis thaliana
Arabidopsis halleri
Camelineae
Turritis glabra
Turritideae
Erysimum aurantiacum
Erysimum sylvestre
Erysimum rhaeticum
Erysimum virgatum
Erysimeae
Olimarabidopsis pumila
Alyssopsideae
Pachycladon cheesemanii
Pachycladon enysii
Microlepidieae
Crucihimalaya wallichii
Crucihimalayeae
Boechera gracilipes
Phoenicaulis cheiranthoides
Polyctenium fremontii
Boechereae
Halimolobos pubens
Halimolobeae
Lesquerella montana
Nerisyrenia camporum
Physarieae
Cardamine flexuosa
Cardamine asarifolia
Cardamine pratensis
Dentaria enneaphyllos
Cardamine resedifolia
Cardamine alpina
Cardamine hirsuta
Cardamine trifolia
Cardamine impatiens
Dentaria pentaphyllos
Cardamine pentaphyllos
Nasturtium officinale
Rorippa sylvestris
Rorippa austriaca
Barbarea verna
Leavenworthia uniflora
Leavenworthia exigua
Cardamineae
Lepidium campestre
Cardaria draba
Lepidium virginicum
Lepidieae
Hutchinsia alpina
Hutchinsia brevicaulis
Hornungia petraea
Hymenolobus pauciflorus
Descurainia bourgaeana
Descurainia sophia
Descurainieae
Smelowskia calycina
Smelowskieae
Brassica napus
Diplotaxis tenuifolia
Brassica repanda susp baldensis
Hirschfeldia incana
Brassiceae
Sisymbrium officinale
Sisymbrieae
Ochthodium aegyptiacum
Unassigned
Stanleya pinnata
Thelypodium laciniatum
Thelypodieae
Isatis tinctoria
Isatideae
Calepina irregularis
Calepineae
Thellungiella halophila
Eutremeae
Peltaria angustifolia
Thlaspideae
Draba aizoides
Draba nemorosa
Draba dubia
Draba verna
Arabis hirsuta Aggregate
Arabis soyeri subsp subcoriacea
Arabis hirsuta
Arabis nova
Arabis alpina
Arabis turrita
Arabideae
Morettia philaeana
Lobularia maritima
Malcolmia littorea
Anastaticeae
Alyssum alissoides
Alyssum dasycarpum
Berteroa incana
Fibigia clypeata
Alysseae
Noccaea praecox
Thlaspi perfoliatum
Noccaea rotundifolium
Thlaspideae
Iberis amara
Iberideae
Heliophila coronopifolia
Cochlearieae
Cochlearia officinalis
Heliophileae
Kernera saxatilis
Kernereae
Biscutella laevigata
Biscutella prealpina
Biscutelleae
Lunaria annua
Unassigned
Bunias orientalis
Buniadeae
Hesperis matronalis
Hesperideae
Matthiola fruticulosa
Anchonieae
Aethionema cordifolium
Aethionema saxatile
Aethionema grandiflorum
Aethionemeae
Cleome spinosa
Cleome hirta
Cleomaceae
Lineage I
Lineage II
Expanded Lineage II
Lineage III
Basal
Outgroup

Figure 3.7: Cladogram of the phylogenetic relationship among 95 Brassicaceae species of 71 CDS based on amino acid substitution. The cladogram represents the consensus topology of BI, MP and ML. Numbers on branches indicate different support values. Black color illustrates Mrbayes support values, blue color represents raxml support values, purple color means phyml support values and green color indicates paup support values.

3.3 Patterns of chloroplast evolution across families in angiosperms

3.3.1 Chloroplast genome assembly

In this part of the study, whether a large scale comparative approach carried out on species that co-evolved in the same floristic assemblage could detect evidences in canalization and recurrence of the molecular signatures associated to plastome evolution was investigated. To this aim, 1,037 individuals from more than 1000 taxa were sequenced by genome skimming through Illumina sequencing. Straub firstly defined "genome skimming" in 2012 as shallow shotgun sequencing of the total genomic DNA (gDNA), which contains the high-copy fraction of the genome (plastome, mitogenome, and repetitive elements) (Straub et al., 2012). The summary of Velvet *de novo* assembly for these species is shown in Table 3.7. The average contig number per taxon was 8937.25, and the means of N50/N90 were 476.46 and 145.82, respectively, indicating that as expected gene sequences from the low-copy number genomic regions were only partially assembled.

Table 3.7: Accessions sampled and assembly summary of 1037 species

UniqueID	Name (Flora of Italy)	Family	Number contigs	Total bases	N50/N90 values	Chloroplast CDS	Mitochondrial CDS	rRNA genes
GS0001	*Stellaria pallida* (Dumort.) Crépin	Caryophyllaceae	3640	1112201	312/152	58	18	3
GS0002	*Hedera helix* L.	Araliaceae	6105	1505587	252/125	76	10	4
GS0004	*Viola riviniana* Rchb.	Violaceae	46	142414	9637/2210	82	7	4
GS0005	*Carex alba* Scop.	Cyperaceae	337	217666	2112/201	64	4	3
GS0006	*Geranium molle* L.	Geraniaceae	21995	5292688	229/139	53	19	2
GS0007	*Cerastium brachypetalum* Desp. ex Pers.	Caryophyllaceae	2842	848777	326/132	5	19	3
GS0008	*Arenaria leptoclados* (Rchb.) Guss.	Caryophyllaceae	7648	2234504	283/128	69	31	3
GS0010	*Euphorbia helioscopia* L. subsp. *helioscopia*	Euphorbiaceae	30235	6948201	225/128	78	37	0
GS0011	*Cerastium semidecandrum* L.	Caryophyllaceae	2758	919485	396/145	53	17	2
GS0012	*Euphorbia cyparissias* L.	Euphorbiaceae	19402	4705694	234/125	69	35	3
GS0013	*Mercurialis annua* L.	Euphorbiaceae	17829	4623303	249/141	71	40	3
GS0015	*Viola alba Besser*	Violaceae	3544	1373840	499/182	68	19	3
GS0016	*Hierochloë australis*	Poaceae	5837	1382569	236/125	70	5	3

	(Schrad.) Roem. et Schult.							
GS0017	*Mercurialis ovata* Sternb. et Hoppe	Euphorbiaceae	11374	3074514	278/129	65	38	3
GS0018	*Carex digitata* L.	Cyperaceae	417	261771	1507/211	63	4	3
GS0019	*Potentilla pusilla Host*	Rosaceae	1681	568274	344/171	77	7	3
GS0020	*Sesleria caerulea* (L.) Ard.	Poaceae	3245	971938	327/141	73	4	3
GS0021	*Erucastrum nasturtiifolium* (Poir.) O. E. Schulz	Brassicaceae	1328	530926	472/175	79	6	4
GS0022	*Schoenus nigricans* L.	Cyperaceae	1822	542000	333/125	28	4	2
GS0023	*Carex flacca Schreb.*	Cyperaceae	1161	438959	467/163	69	4	2
GS0024	*Stellaria media* (L.) Vill.	Caryophyllaceae	7740	1905161	242/127	58	20	3
GS0025	*Geranium rotundifolium* L.	Geraniaceae	5092	1305402	254/125	55	14	2
GS0026	*Lamium purpureum* L.	Lamiaceae	3321	927804	333/125	68	6	4
GS0027	*Viola collina Besser*	Violaceae	2351	773464	325/193	76	9	3
GS0028	*Veronica persica* Poir.	Plantaginaceae	2013	623547	331/140	71	7	3
GS0030	*Poa annua* L.	Poaceae	7902	1862017	235/125	75	4	3
GS0031	*Glechoma hederacea* L.	Lamiaceae	3605	895207	244/125	78	8	3
GS0032	*Senecio vulgaris* L.	Asteraceae	8720	2364862	284/132	79	28	3
GS0034	*Alopecurus myosuroides* Huds.	Poaceae	13904	2962060	212/125	77	5	3
GS0036	*Erodium cicutarium* (L.) L'Hér.	Geraniaceae	4398	1102081	253/125	71	5	2
GS0037	*Carex acuta* L.	Cyperaceae	2026	619495	295/167	64	4	2
GS0038	*Veronica hederifolia* L.	Plantaginaceae	413	205870	2532/154	68	6	3
GS0039	*Poa trivialis* L.	Poaceae	2949	815708	289/125	77	3	4
GS0040	*Rhamnus pumila Turra*	Rhamnaceae	36688	9002150	230/189	69	42	4
GS0042	*Centaurea scabiosa* L.	Asteraceae	11059	2955228	275/133	78	35	4
GS0043	*Viola arvensis Murray*	Violaceae	1549	504213	336/147	76	7	3
GS0044	*Alliaria petiolata* (M. Bieb.) Cavara et Grande	Brassicaceae	2622	901094	409/154	74	11	2
GS0047	*Chelidonium majus* L.	Papaveraceae	5477	1568745	309/128	72	16	2
GS0048	*Poa angustifolia* L.	Poaceae	10992	2670343	250/125	78	4	3
GS0049	*Vicia sativa* L.	Fabaceae	15646	3352560	214/125	72	19	3
GS0050	*Fumaria officinalis* L.	Papaveraceae	9498	2845457	265/184	46	41	4
GS0051	*Muscari comosum* (L.) Mill.	Hyacinthaceae	26239	5274651	201/125	69	30	3
GS0052	*Carduus pycnocephalus* L.	Asteraceae	18496	4219606	227/125	80	35	3
GS0053	*Ranunculus bulbosus* L.	Ranunculaceae	27855	5433248	196/125	76	6	3
GS0054	*Viola reichenbachiana* Jord. ex Boreau	Violaceae	8353	2721187	364/145	70	35	3
GS0055	*Equisetum arvense* L.	Equisetaceae	4292	1262211	317/126	82	14	2
GS0056	*Hordeum murinum* L.	Poaceae	6889	1294588	170/125	74	3	3
GS0057	*Ulmus glabra* Huds.	Ulmaceae	6566	1986205	350/132	58	31	3
GS0058	*Lathyrus vernus* (L.) Bernh.	Fabaceae	10583	2796505	278/135	56	11	3
GS0059	*Primula vulgaris* Huds. subsp. *vulgaris*	Primulaceae	3331	1138453	453/128	77	38	3
GS0060	*Anemone trifolia* L.	Ranunculaceae	12671	2872209	232/125	33	8	4
GS0061	*Mercurialis ovata* Sternb. et Hoppe	Euphorbiaceae	8290	2364345	310/129	68	34	3

GS0062	*Daphne mezereum* L.	Thymelaeaceae	7396	1920729	261/126	50	20	3
GS0063	*Viola hirta* L.	Violaceae	1706	573470	328/190	73	7	3
GS0064	*Gagea villosa* (M. Bieb.) Sweet	Liliaceae	2256	682005	310/143	66	5	1
GS0065	*Taraxacum officinale (aggregatum)*	Asteraceae	10980	2913003	273/134	76	25	3
GS0066	*Anisantha sterilis (*L.*)* Nevski	Poaceae	11338	2246753	194/125	77	4	4
GS0068	*Medicago lupulina* L.	Fabaceae	15939	3490682	219/125	63	14	2
GS0069	*Viola odorata* L.	Violaceae	4727	1858412	503/157	74	40	3
GS0070	*Symphytum tuberosum* L. subsp. *angustifolium* (A. Kern.) Nyman	Boraginaceae	241	193742	3752/253	65	6	2
GS0071	*Phyllitis scolopendrium* (L.) Newman subsp. *scolopendrium*	Aspleniaceae	4524	1384990	356/140	0	1	2
GS0072	*Taxus baccata* L.	Taxaceae	152	178279	4158/372	74	2	3
GS0073	*Corydalis cava* (L.) Schweigg. et Körte subsp. *cava*	Papaveraceae	1642	569805	381/155	16	5	3
GS0074	*Ficaria verna Hudson*	Ranunculaceae	9672	1944861	197/125	71	5	3
GS0076	*Viola alba Besser*	Violaceae	4243	1645037	518/170	73	25	3
GS0077	*Polygala chamaebuxus* L.	Polygalaceae	642	227335	359/197	23	1	3
GS0078	*Carex praecox* Schreb.	Cyperaceae	2404	711353	284/173	73	5	2
GS0079	*Cerastium glutinosum* Fr.	Caryophyllaceae	12227	2756201	224/125	55	26	4
GS0080	*Bellis perennis* L.	Asteraceae	15083	3000607	193/125	67	17	3
GS0081	*Melica nutans* L.	Poaceae	11426	2697252	242/125	70	9	3
GS0082	*Prunus mahaleb* L.	Rosaceae	1946	944802	980/157	78	28	4
GS0083	*Emerus major* Mill.	Fabaceae	7920	1977298	245/125	74	30	3
GS0084	*Amelanchier ovalis* Medik.	Rosaceae	1425	567595	469/179	75	8	1
GS0085	*Vincetoxicum hirundinaria* Medik.	Apocynaceae	2182	1116087	1004/193	58	37	3
GS0086	*Ostrya carpinifolia* Scop.	Betulaceae	1230	471675	354/201	70	9	2
GS0087	*Globularia cordifolia* L.	Plantaginaceae	3108	1173894	498/167	60	18	3
GS0088	*Globularia bisnagarica* L.	Plantaginaceae	2989	1232483	567/188	62	18	3
GS0089	*Carex michelii Host*	Cyperaceae	1449	504963	336/192	72	4	1
GS0091	*Adiantum capillus-veneris* L.	Pteridaceae	2404	959252	655/146	79	1	4
GS0093	*Ruscus aculeatus* L.	Asparagaceae	2878	810903	285/132	72	5	1
GS0094	*Helianthemum apenninum* (L.) Mill. subsp. *apenninum*	Cistaceae	21780	4238686	195/125	21	5	2
GS0095	*Veronica arvensis* L.	Plantaginaceae	1681	750504	1847/138	78	33	4
GS0096	*Cerastium glomeratum* Thuill.	Caryophyllaceae	2296	670361	313/127	62	12	3
GS0097	*Valerianella locusta* (L.) Laterr.	Valerianaceae	112298	30386583	263/193	61	29	0
GS0098	*Asplenium adiantum-nigrum* L. subsp. *adiantum-nigrum*	Aspleniaceae	9779	2474661	265/125	17	3	1
GS0099	*Saxifraga tridactylites* L.	Saxifragaceae	2931	1052026	500/145	48	26	3
GS0100	*Glechoma hirsuta* Waldst. et Kit.	Lamiaceae	4739	1330547	308/125	73	13	3
GS0102	*Asplenium trichomanes* L.	Aspleniaceae	9111	2458209	289/126	1	1	2

GS0103	*Viburnum lantana* L.	Adoxaceae	10973	2567480	236/127	29	11	3
GS0104	*Carex humilis* Leyss.	Cyperaceae	1490	517790	426/151	66	3	2
GS0105	*Scorzonera austriaca* Willd.	Asteraceae	5771	1556927	277/139	79	6	4
GS0106	*Crepis froelichiana* DC.	Asteraceae	22105	4585884	207/125	70	7	2
GS0107	*Leontodon incanus* (L.) Schrank	Asteraceae	35176	7678719	218/125	77	29	2
GS0108	*Carex halleriana Asso*	Cyperaceae	1528	560976	464/166	67	4	2
GS0109	*Buglossoides purpurocaerulea* (L.) I. M. Johnst.	Boraginaceae	2542	791066	299/187	57	9	3
GS0110	*Cytisus purpureus* Scop.	Fabaceae	7092	2204596	362/142	73	19	3
GS0111	*Vicia hirsuta* (L.) Gray	Fabaceae	4043	1141007	298/133	41	6	3
GS0112	*Saponaria ocymoides* L.	Caryophyllaceae	4499	1076567	235/125	41	5	4
GS0113	*Helianthemum canum* (L.) Baumg.	Cistaceae	7273	1629167	224/126	23	4	1
GS0114	*Viscum album L.*	Viscaceae	2437	667312	286/125	60	6	3
GS0116	*Muscari neglectum* Guss. ex Ten.	Hyacinthaceae	22596	4550700	203/125	73	15	3
GS0118	*Ajuga genevensis* L.	Lamiaceae	2973	821463	270/134	78	6	3
GS0119	*Thymus praecox* Opiz	Lamiaceae	4764	1379428	320/125	58	23	2
GS0120	*Carex liparocarpos Gaudin*	Cyperaceae	2110	658873	311/163	67	4	2
GS0121	*Cymbalaria muralis* G. Gaertn., B. Mey. et Scherb	Plantaginaceae	2561	1024420	559/172	67	16	3
GS0122	*Crepis vesicaria* L.	Asteraceae	14437	2901036	200/125	76	10	4
GS0123	*Tussilago farfara* L.	Asteraceae	7665	1969444	262/125	76	26	3
GS0124	*Viola rupestris* A. F. W. Schmidt subsp. *rupestris*	Violaceae	3292	1328799	632/157	77	36	4
GS0125	*Pulsatilla montana* (Hoppe) Rchb.	Ranunculaceae	16887	4019904	245/125	63	17	4
GS0126	*Epimedium alpinum* L.	Berberidaceae	4057	1112908	270/140	77	5	3
GS0127	*Pulmonaria angustifolia* L.	Boraginaceae	3239	1050442	389/135	56	27	4
GS0128	*Viola mirabilis* L.	Violaceae	3315	1351628	634/167	72	23	3
GS0129	*Carex montana* L.	Cyperaceae	575	283938	872/195	71	4	3
GS0130	*Colchicum autumnale* L.	Colchicaceae	7357	1852082	259/126	3	6	1
GS0131	*Lilium martagon* L.	Liliaceae	1382	492990	418/148	76	5	1
GS0132	*Primula veris* L.	Primulaceae	2568	910203	515/141	76	24	3
GS0133	*Carex tomentosa* L.	Cyperaceae	207	184361	4246/243	67	4	2
GS0134	*Carex panicea* L.	Cyperaceae	1255	477333	446/180	61	3	2
GS0135	*Primula veris* L.	Primulaceae	2213	793185	472/147	76	17	3
GS0136	*Carex caryophyllea Latourr.*	Cyperaceae	1343	489138	454/159	66	4	3
GS0137	*Lamium album* L. subsp. *album*	Lamiaceae	7476	1783445	242/125	71	7	2
GS0138	*Rumex acetosa* L.	Polygonaceae	2228	594020	261/129	68	6	3
GS0140	*Ornithogalum umbellatum* L.	Hyacinthaceae	20204	4407626	221/125	70	5	2
GS0142	*Carex davalliana* Sm.	Cyperaceae	4385	1032370	231/125	72	4	2
GS0143	*Fragaria vesca* L. subsp. *vesca*	Rosaceae	1320	558650	491/203	79	12	2
GS0145	*Crocus albiflorus* Kit.	Iridaceae	6948	1579324	225/125	81	5	2
GS0146	*Orchis purpurea* Huds.	Orchidaceae	21648	4772798	225/125	71	12	3
GS0147	*Veratrum nigrum* L.	Melanthiaceae	8383	2371313	310/137	76	15	2

GS0148	*Draba nemorosa* L.	Brassicaceae	1341	471353	359/167	78	8	2
GS0151	*Silene dioica* (L.) Clairv.	Caryophyllaceae	7133	1638730	230/125	75	5	3
GS0152	*Arabidopsis halleri* (L.) O'Kane et Al-Shehbaz	Brassicaceae	2484	971874	588/154	75	27	3
GS0153	*Viola tricolor* L.	Violaceae	6331	1582585	255/127	77	7	3
GS0155	*Anthoxanthum odoratum* L.	Poaceae	18940	3727451	199/125	75	4	4
GS0156	*Potentilla micrantha* Ramond ex DC.	Rosaceae	3316	982280	303/152	75	10	3
GS0157	*Rumex arifolius* All.	Polygonaceae	5107	1200943	233/125	77	6	4
GS0158	*Luzula campestris* (L.) DC.	Juncaceae	4597	1003501	208/125	5	2	3
GS0159	*Anthriscus sylvestris* (L.) Hoffm. subsp. *sylvestris*	Apiaceae	10823	2341415	213/125	77	6	3
GS0160	*Dactylis glomerata* L.	Poaceae	17097	3200056	185/125	76	3	3
GS0161	*Plantago lanceolata* L.	Plantaginaceae	7473	1754377	232/125	74	5	3
GS0162	*Cerastium lucorum Schur*	Caryophyllaceae	4828	1221505	255/125	46	9	2
GS0163	*Cardamine amara* L.	Brassicaceae	2507	795953	324/157	78	9	2
GS0164	*Chrysosplenium alternifolium* L.	Saxifragaceae	1818	671350	477/148	39	7	3
GS0165	*Ajuga reptans* L.	Lamiaceae	4830	1066815	216/125	75	6	3
GS0169	*Chenopodium bonus-henricus* L.	Amaranthaceae	4011	1281583	438/125	74	30	3
GS0170	*Luzula pilosa* (L.) Willd.	Juncaceae	10610	2501985	232/127	11	14	2
GS0171	*Corydalis intermedia* (L.) Mérat	Papaveraceae	8860	2296385	259/131	29	5	3
GS0172	*Salvia pratensis* L.	Lamiaceae	12029	3107232	259/128	77	33	2
GS0173	*Crepis biennis* L.	Asteraceae	39765	8109275	205/125	57	33	2
GS0174	*Arrhenatherum elatius* (L.) P. Beauv. ex J. et C. Presl	Poaceae	20891	3841177	176/125	77	5	4
GS0175	*Bromus hordeaceus* L.	Poaceae	15873	2918870	175/125	76	4	3
GS0177	*Cystopteris fragilis* (L.) Bernh.	Cystopteridaceae	6624	2008190	346/140	9	1	3
GS0178	*Adoxa moschatellina* L.	Adoxaceae	2354	736466	344/140	61	6	3
GS0179	*Viola thomasiana* Songeon et Perr.	Violaceae	8813	2636654	331/131	70	35	3
GS0180	*Ajuga pyramidalis* L.	Lamiaceae	6440	1496236	232/125	78	6	3
GS0181	*Cytisus hirsutus* L.	Fabaceae	8744	2278761	275/131	74	13	2
GS0182	*Petasites albus* (L.) *Gaertn.*	Asteraceae	19006	4593421	248/127	79	15	3
GS0183	*Oxalis acetosella* L.	Oxalidaceae	19771	5037884	256/137	62	14	3
GS0184	*Pulmonaria officinalis* L.	Boraginaceae	6025	1439551	237/126	53	10	4
GS0185	*Pteridium aquilinum* (L.) *Kuhn* subsp. *aquilinum*	Dennstaedtiaceae	4506	1360097	351/130	85	2	3
GS0186	*Dactylorhiza sambucina* (L.) Soó	Orchidaceae	15634	3899890	255/141	78	9	3
GS0187	*Atocion rupestre* (L.) *Rafin.*	Caryophyllaceae	4189	1169528	293/135	40	6	4
GS0188	*Potentilla rupestris* L.	Rosaceae	10537	2931578	247/188	74	36	4
GS0189	*Polygonatum odoratum* (Mill.) Druce	Asparagaceae	17412	3694900	216/125	77	5	2
GS0191	*Lathraea squamaria* L.	Orobanchaceae	31740	6243342	195/125	39	37	1
GS0192	*Vicia sepium* L.	Fabaceae	9503	2075063	218/125	71	8	3
GS0193	*Polygonatum multiflorum* (L.) All.	Asparagaceae	19226	4677727	250/125	79	22	1
GS0194	*Gagea lutea* (L.) Ker	Liliaceae	6130	1517700	254/126	69	5	1

	Gawl.							
GS0198	*Matricaria discoidea* DC.	Asteraceae	8691	1946358	223/125	76	12	3
GS0199	*Veronica serpyllifolia* L.	Plantaginaceae	3106	933666	385/125	70	19	3
GS0200	*Malva neglecta* Wallr.	Malvaceae	9420	2316021	239/125	79	20	3
GS0201	*Lotus corniculatus* L.	Fabaceae	18337	4272767	232/125	74	21	3
GS0202	*Bromopsis inermis* (Leysser) Holub	Poaceae	20462	4031134	200/125	77	4	2
GS0206	*Geranium columbinum* L.	Geraniaceae	4878	1111152	223/125	36	4	2
GS0207	*Trifolium scabrum* L. subsp. *scabrum*	Fabaceae	29295	6894789	226/132	66	31	3
GS0208	*Galium album Miller*	Rubiaceae	2041	765683	484/163	78	6	0
GS0209	*Bromopsis erecta* (Hudson) Fourr.	Poaceae	8696	1859388	214/125	74	5	3
GS0210	*Onobrychis viciifolia* Scop.	Fabaceae	10251	2876756	316/126	56	21	2
GS0211	*Tragopogon dubius* Scop.	Asteraceae	32104	6834349	212/125	76	32	3
GS0213	*Cornus sanguinea* L.	Cornaceae	1461	496646	372/144	47	6	3
GS0215	*Euonymus europaeus* L.	Celastraceae	12754	4023907	329/125	73	40	2
GS0216	*Vicia cracca* L.	Fabaceae	5707	1474973	266/132	51	6	3
GS0217	*Melittis melissophyllum* L.	Lamiaceae	14187	3254375	221/125	78	34	2
GS0218	*Cruciata glabra* (L.) Ehrend.	Rubiaceae	9115	2567086	257/188	57	33	3
GS0219	*Corylus avellana* L.	Betulaceae	1735	638118	407/196	75	9	2
GS0220	*Polygala comosa Schkuhr*	Polygalaceae	2873	864081	308/159	31	7	4
GS0221	*Hippocrepis comosa* L. subsp. *comosa*	Fabaceae	19570	4460479	230/125	70	30	3
GS0222	*Trinia glauca* (L.) Dumort.	Apiaceae	8748	2162652	264/125	73	12	4
GS0224	*Trifolium montanum* L.	Fabaceae	6003	1638395	299/134	63	6	3
GS0225	*Silene nutans* L.	Caryophyllaceae	14429	3113996	219/125	66	6	3
GS0226	*Hieracium bifidum* Kit. ex Hornem.	Asteraceae	18255	4108297	227/125	77	11	4
GS0227	*Genista germanica* L.	Fabaceae	7944	2578790	416/131	75	37	2
GS0228	*Rhamnus saxatilis* Jacq.	Rhamnaceae	28361	6584912	223/125	58	41	3
GS0229	*Ranunculus acris* L.	Ranunculaceae	33638	6713452	200/125	79	11	4
GS0231	*Persicaria amphibia* (L.) Delarbre	Polygonaceae	4384	1239085	297/134	71	12	3
GS0232	*Carex hirta* L.	Cyperaceae	3128	982729	300/194	66	6	2
GS0233	*Alopecurus pratensis* L.	Poaceae	8239	1950216	241/125	75	3	2
GS0234	*Equisetum palustre* L.	Equisetaceae	9160	2226288	242/125	81	13	3
GS0236	*Lamium orvala* L.	Lamiaceae	3779	1283635	433/141	78	18	2
GS0237	*Glechoma hirsuta* Waldst. et Kit.	Lamiaceae	4619	1192643	270/125	77	8	4
GS0239	*Euphorbia dulcis* L.	Euphorbiaceae	12315	2787667	229/125	44	6	3
GS0240	*Euphorbia carniolica* Jacq.	Euphorbiaceae	25832	5933024	236/125	36	16	2
GS0242	*Ranunculus tuberosus* Lapeyr.	Ranunculaceae	44980	8673134	193/125	77	18	3
GS0243	*Maianthemum bifolium* (L.) Schmidt	Asparagaceae	3349	1032969	331/150	73	5	2
GS0244	*Actaea spicata* L.	Ranunculaceae	177	201168	38088/312	76	6	3
GS0245	*Salix caprea* L.	Salicaceae	5850	1721667	338/127	71	15	4
GS0246	*Paris quadrifolia* L.	Melanthiaceae	1397	509616	414/155	75	5	3
GS0247	*Alnus incana* (L.) Moench	Betulaceae	4120	1254981	317/158	68	11	3

GS0248	*Lonicera xylosteum* L.	Caprifoliaceae	28970	6967446	230/125	69	42	2
GS0249	*Valeriana tripteris* L.	Valerianaceae	10008	3135255	443/126	55	21	3
GS0250	*Galium odoratum* (L.) Scop.	Rubiaceae	4601	1418878	305/159	57	28	3
GS0251	*Philadelphus coronarius* L.	Hydrangeaceae	4762	1760373	556/132	58	41	3
GS0252	*Mercurialis perennis* L.	Euphorbiaceae	12034	3162545	272/126	65	38	4
GS0253	*Polystichum aculeatum* (L.) Roth	Dryopteridaceae	21122	5582504	288/125	78	1	3
GS0254	*Ranunculus carinthiacus Hoppe*	Ranunculaceae	37047	7799794	211/125	76	17	3
GS0255	*Moehringia trinervia* (L.) Clairv.	Caryophyllaceae	6116	1713210	311/128	51	23	3
GS0256	*Ornithogalum kochii* Parl.	Hyacinthaceae	25673	5379056	212/125	80	4	2
GS0257	*Carum carvi* L. subsp. *carvi*	Apiaceae	16865	3692324	224/125	76	12	3
GS0259	*Viola canina* L.	Violaceae	3217	1363811	584/192	78	35	3
GS0260	*Alchemilla exigua* Buser ex Paulin	Rosaceae	2039	622301	326/134	83	6	4
GS0261	*Polygala alpestris* Rchb.	Polygalaceae	3406	1004498	298/158	10	7	4
GS0262	*Poa supina* Schrad.	Poaceae	60762	13176217	219/125	32	23	4
GS0263	*Vicia oroboides* Wulfen	Fabaceae	13686	2885722	211/125	68	9	3
GS0265	*Ranunculus lanuginosus* L.	Ranunculaceae	9905	2194808	219/125	78	6	4
GS0267	*Veronica peregrina* L. subsp. *peregrina*	Plantaginaceae	3536	1116414	382/138	79	22	4
GS0268	*Rumex pulcher* L.	Polygonaceae	9879	2537497	260/133	76	11	3
GS0269	*Capsella rubella* Reut.	Brassicaceae	4971	1123115	219/125	77	11	3
GS0270	*Avena barbata* Pott ex Link	Poaceae	17497	3581337	211/125	41	3	2
GS0271	*Festuca rubra* L.	Poaceae	29957	6229703	211/125	76	5	4
GS0272	*Stachys recta* L.	Lamiaceae	5078	1483642	342/125	77	19	2
GS0273	*Cruciata laevipes* Opiz	Rubiaceae	7874	2268798	278/170	35	21	3
GS0274	*Brachypodium rupestre* (Host) Roem. et Schult.	Poaceae	14571	3690979	241/132	71	26	4
GS0275	*Centranthus ruber* (L.) *DC.* subsp. *ruber*	Valerianaceae	11680	3848304	355/134	59	31	3
GS0276	*Geranium pyrenaicum* Burm. fil. subsp. *pyrenaicum*	Geraniaceae	5494	1166958	199/125	56	4	1
GS0277	*Sisymbrium officinale* (L.) Scop.	Brassicaceae	5886	1611317	274/144	77	15	4
GS0278	*Carex divulsa Stokes*	Cyperaceae	428	218399	792/206	57	3	2
GS0279	*Parietaria judaica* L.	Urticaceae	9386	2511165	260/125	72	29	1
GS0280	*Medicago sativa* L.	Fabaceae	6164	1430322	236/125	75	8	3
GS0281	*Papaver rhoeas* L.	Papaveraceae	4533	1136735	246/125	78	6	3
GS0282	*Medicago minima* (L.) L.	Fabaceae	8869	2008754	229/125	65	17	3
GS0283	*Malus sylvestris* (L.) Mill.	Rosaceae	2464	777656	319/155	80	8	0
GS0284	*Rumex crispus* L.	Polygonaceae	8503	2146019	256/131	76	9	2
GS0285	*Sedum dasyphyllum* L.	Crassulaceae	1611	689647	562/191	64	23	2
GS0286	*Geranium robertianum* L.	Geraniaceae	5903	1564619	291/125	49	8	1
GS0287	*Silene nutans* L.	Caryophyllaceae	29184	6449838	228/125	70	9	3
GS0288	*Tamus communis* L.	Dioscoreaceae	14202	3743946	246/132	69	42	1
GS0289	*Orobanche hederae* Duby	Orobanchaceae	7146	2924186	900/142	30	29	3

GS0290	*Convolvulus cantabrica* L.	Convolvulaceae	6324	1686741	284/128	70	24	4
GS0291	*Medicago orbicularis* (L.) Bartal.	Fabaceae	6040	1596489	257/125	68	26	3
GS0292	*Pistacia terebinthus* L. subsp. *terebinthus*	Terebinthaceae	11966	3438148	267/125	66	35	3
GS0293	*Avena sterilis* L.	Poaceae	10929	2082428	185/125	78	4	3
GS0295	*Lamiastrum galeobdolon* L.	Lamiaceae	4755	1268921	277/128	82	11	3
GS0297	*Aristolochia clematitis* L.	Aristolochiaceae	719	598866	3254/219	29	42	3
GS0298	*Campanula carnica* Mert. et W. D. J. Koch	Campanulaceae	12406	3145844	260/128	74	25	3
GS0299	*Chaerophyllum aureum* L.	Apiaceae	14085	3288759	237/125	78	11	2
GS0300	*Scandix pecten-veneris* L.	Apiaceae	6250	1671634	280/125	74	23	3
GS0301	*Piptatherum miliaceum* (L.) Coss.	Poaceae	21625	4918071	218/129	75	26	3
GS0302	*Geranium purpureum* Vill.	Geraniaceae	71	156620	17045/829	71	5	2
GS0303	*Polypodium cambricum* L.	Polypodiaceae	3880	1238661	384/135	5	1	2
GS0304	*Buxus sempervirens* L.	Buxaceae	3928	1809704	1013/171	63	30	4
GS0305	*Helleborus foetidus* L.	Ranunculaceae	10278	2106989	204/125	42	12	4
GS0306	*Quercus ilex* L. subsp. *ilex*	Fagaceae	8623	2520129	274/138	78	42	3
GS0307	*Pseudofumaria lutea* (L.) Borkh.	Papaveraceae	7152	2155697	289/155	35	32	3
GS0308	*Sonchus asper* (L.) Hill	Asteraceae	11484	2716332	230/125	76	34	3
GS0309	*Sonchus oleraceus* L.	Asteraceae	14342	3163473	212/125	77	29	4
GS0310	*Argyrolobium zanonii* (Turra) P. W. Ball subsp. zanonii	Fabaceae	12265	3079487	255/126	73	24	4
GS0311	*Dictamnus albus* L.	Rutaceae	13228	3125368	241/125	74	13	3
GS0313	*Bromopsis condensata* (Hackel) Holub	Poaceae	21575	4379936	206/125	75	6	4
GS0314	*Arctostaphylos uva-ursi* (L.) Spreng.	Ericaceae	1045	414574	494/180	60	6	2
GS0315	*Neottia nidus-avis* (L.) Rich.	Orchidaceae	3624	871850	239/127	22	6	3
GS0316	*Euphorbia amygdaloides* L.	Euphorbiaceae	18598	3983604	219/125	54	7	4
GS0317	*Laburnum anagyroides* Medik.	Fabaceae	4336	1774445	644/164	76	29	3
GS0318	*Campanula rotundifolia* L. subsp. *rotundifolia*	Campanulaceae	9807	2284659	235/125	63	5	3
GS0319	*Thesium bavarum Schrank*	Thesiaceae	12639	3549390	270/151	58	36	1
GS0322	*Leucanthemum ircutianum* Turcz. ex DC.	Asteraceae	7219	1625198	221/125	82	5	3
GS0323	*Luzula forsteri* (Sm.) DC.	Juncaceae	1546	369248	248/125	18	1	3
GS0324	*Dryopteris carthusiana* (Vill.) H. P. Fuchs	Dryopteridaceae	10737	2958666	296/130	35	1	3
GS0325	*Aquilegia atrata* W. D. J. Koch	Ranunculaceae	15292	4005813	248/128	67	39	1
GS0326	*Anemone nemorosa* L.	Ranunculaceae	3667	944947	256/133	69	6	2
GS0327	*Rhinanthus alectorolophus* (Scop.) Pollich	Orobanchaceae	9916	2440548	252/125	39	27	3
GS0328	*Asarum europaeum* L.	Aristolochiaceae	16838	3965740	241/125	42	15	3

GS0329	*Ranunculus venetus* Landolt	Ranunculaceae	8716	2109150	248/126	78	6	3
GS0330	*Veronica prostrata* L. subsp. *prostrata*	Plantaginaceae	13193	3328663	261/127	70	23	3
GS0331	*Cephalanthera longifolia* (L.) Fritsch	Orchidaceae	18400	4123556	231/125	66	10	3
GS0332	*Thalictrum minus* L.	Ranunculaceae	15452	3915340	247/125	80	33	3
GS0334	*Geranium phaeum* L.	Geraniaceae	8905	1969829	221/125	55	6	2
GS0335	*Vicia loiseleurii* (M. Bieb.) Litv.	Fabaceae	4487	1134443	257/126	36	5	3
GS0336	*Scrophularia canina* L.	Scrophulariaceae	346	225947	14612/212	80	5	3
GS0338	*Anisantha diandra* (Roth) Tutin ex Tzvelev	Poaceae	8098	1633574	199/125	77	4	3
GS0339	*Chaenorhinum minus* (L.) Lange	Plantaginaceae	972	743791	1944/250	72	36	2
GS0340	*Minuartia hybrida* (Vill.) Shischk.	Caryophyllaceae	2962	956970	318/190	67	7	2
GS0341	*Salvia verbenaca* L.	Lamiaceae	17060	4042438	237/125	77	28	2
GS0342	*Helminthotheca echioides* (L.) Holub	Asteraceae	14142	3274552	226/125	79	35	4
GS0343	*Calendula arvensis* (Vaill.) L.	Asteraceae	20339	4095779	199/125	78	16	2
GS0344	*Anisantha madritensis* (L.) Nevski	Poaceae	6695	1539199	231/125	79	4	3
GS0345	*Vicia sativa* L.	Fabaceae	20241	4303046	210/125	74	25	3
GS0346	*Verbascum blattaria* L.	Scrophulariaceae	7806	2369541	274/158	69	38	4
GS0351	*Cirsium vulgare* (Savi) Ten.	Asteraceae	20378	4863054	242/125	80	27	3
GS0352	*Rostraria cristata* (L.) Tzvelev	Poaceae	3264	879513	277/132	72	4	3
GS0353	*Aphanes arvensis* L.	Rosaceae	975	304827	362/136	66	5	3
GS0354	*Herniaria hirsuta* L.	Caryophyllaceae	4377	1111870	260/125	49	9	3
GS0355	*Euphorbia peplus* L.	Euphorbiaceae	8138	2524764	335/127	68	25	3
GS0356	*Crepis capillaris* (L.) Wallr.	Asteraceae	6898	1652041	241/125	80	11	3
GS0357	*Silene alba* (Miller) Krause	Caryophyllaceae	6437	1530749	237/125	76	5	3
GS0358	*Erigeron annuus* (L.) Desf.	Asteraceae	12788	3057037	244/127	69	17	3
GS0359	*Polycarpon tetraphyllum* (L.) L.	Caryophyllaceae	3583	1145716	365/150	38	17	2
GS0360	*Potentilla reptans* L.	Rosaceae	2140	650700	317/127	77	8	4
GS0361	*Barbarea vulgaris* R. Br.	Brassicaceae	4783	1324292	290/129	75	16	3
GS0362	*Phalaroides arundinacea* (L.) Rauschert	Poaceae	3020	918314	328/140	75	5	3
GS0363	*Salix purpurea* L.	Salicaceae	9954	2849469	311/130	72	18	3
GS0364	*Salix eleagnos* Scop.	Salicaceae	8216	2432869	332/125	71	30	4
GS0365	*Populus nigra* L.	Salicaceae	1610	542355	313/192	75	6	2
GS0366	*Rubus caesius* L.	Rosaceae	2527	820628	358/151	74	10	4
GS0367	*Salix alba* L.	Salicaceae	1356	428219	314/125	67	7	3
GS0368	*Schedonorus arundinaceus* (Schreb.) Dumort.	Poaceae	2499	803533	364/161	73	5	4
GS0370	*Diplotaxis muralis* (L.) DC.	Brassicaceae	1757	675966	495/168	76	12	3
GS0371	*Equisetum hyemale* L.	Equisetaceae	7453	1645350	220/125	83	4	2
GS0372	*Melica ciliata* L.	Poaceae	9567	2368671	254/125	76	6	3

GS0373	*Moehringia muscosa* L.	Caryophyllaceae	5231	1371924	271/125	30	13	2
GS0374	*Leontodon crispus* Vill. subsp. *crispus*	Asteraceae	12930	3246577	256/125	82	26	4
GS0375	*Solanum dulcamara* L.	Solanaceae	1964	960367	808/196	75	26	2
GS0376	*Geum urbanum* L.	Rosaceae	1897	539936	312/130	62	6	2
GS0377	*Ranunculus repens* L.	Ranunculaceae	6503	1569582	246/130	75	6	3
GS0378	*Carpinus betulus* L.	Betulaceae	465	209509	620/196	57	4	3
GS0379	*Phyteuma scheuchzeri* All.	Campanulaceae	8172	2201135	291/127	71	19	3
GS0380	*Adiantum capillus-veneris* L.	Pteridaceae	1738	751115	812/162	77	2	3
GS0381	*Limodorum abortivum* (L.) Sw.	Orchidaceae	6207	2297920	505/163	58	14	2
GS0383	*Rumex scutatus* L.	Polygonaceae	6518	1562949	239/125	71	7	2
GS0385	*Asplenium ruta-muraria* L.	Aspleniaceae	20822	4036424	177/125	35	0	3
GS0386	*Cercis siliquastrum* L. subsp. *siliquastrum*	Fabaceae	43794	10623398	232/164	73	40	2
GS0387	*Hieracium piloselloides* Vill.	Asteraceae	6875	1719316	261/130	74	13	1
GS0388	*Ptychotis saxifraga* (L.) Loret et Barrandon	Apiaceae	16615	3438972	201/125	76	23	3
GS0389	*Crepis pulchra* L. subsp. *pulchra*	Asteraceae	8563	1973836	231/125	80	12	2
GS0390	*Ligustrum vulgare* L.	Oleaceae	6315	2270411	462/164	75	23	3
GS0391	*Fumana procumbens* (Dunal) Gren. et Godr.	Cistaceae	11086	2868085	259/125	15	29	2
GS0392	*Trifolium campestre* Schreb.	Fabaceae	22908	5654440	237/143	72	31	3
GS0393	*Catapodium rigidum* (L.) C. E. Hubb.	Poaceae	12709	2534881	198/125	76	4	3
GS0394	*Lapsana communis* L. subsp. *communis*	Asteraceae	6263	1545633	251/125	78	11	3
GS0395	*Papaver dubium* L.	Papaveraceae	12567	2973420	242/125	79	21	3
GS0396	*Cytisophyllum sessilifolium* (L.) O. Lang	Fabaceae	3415	1476142	730/161	75	33	3
GS0397	*Euphorbia esula* L.	Euphorbiaceae	9684	2553613	279/125	71	21	3
GS0398	*Sorbus torminalis* (L.) Crantz	Rosaceae	22952	5278759	222/125	74	39	4
GS0399	*Quercus pubescens* Willd. subsp. *pubescens*	Fagaceae	2987	1033934	469/136	73	28	2
GS0400	*Fraxinus ornus* L. subsp. *ornus*	Oleaceae	3712	1151782	311/176	78	7	4
GS0401	*Rubus canescens* DC.	Rosaceae	16314	4263329	256/135	55	32	3
GS0402	*Tragopogon pratensis* L.	Asteraceae	10867	2569751	239/125	76	9	3
GS0403	*Cyanus triumfetti* (All.) Dostál ex Á. et D. Löve	Asteraceae	16187	3982373	251/127	78	27	2
GS0404	*Seseli annuum* L.	Apiaceae	6095	1708330	305/136	75	8	4
GS0405	*Thesium linophyllon* L.	Thesiaceae	4919	1593540	325/165	47	36	3
GS0406	*Acer campestre* L.	Aceraceae	3150	1629654	1616/168	76	39	3
GS0407	*Echium vulgare* L.	Boraginaceae	2695	865380	353/150	62	18	4
GS0408	*Filipendula vulgaris* Moench	Rosaceae	5841	1817223	326/174	56	8	3
GS0409	*Larix decidua* Mill.	Pinaceae	3188	875260	286/125	71	4	3
GS0410	*Plantago media* L.	Plantaginaceae	1500	404227	279/128	71	3	4
GS0411	*Homalotrichon pubescens* (Huds.) Banfi, Galasso et Bracchi	Poaceae	8913	2058197	233/125	75	3	3

GS0412	*Aster alpinus* L. subsp. *alpinus*	Asteraceae	37388	7967673	214/125	69	22	3
GS0413	*Trollius europaeus* L.	Ranunculaceae	10817	2440640	226/125	79	4	3
GS0414	*Convallaria majalis* L.	Asparagaceae	6092	1615031	272/136	67	5	1
GS0415	*Valeriana dioica* L.	Valerianaceae	17118	4116682	257/125	42	18	4
GS0416	*Gentiana acaulis* L.	Gentianaceae	23114	4843016	208/125	70	27	3
GS0417	*Carex ornithopoda* Willd.	Cyperaceae	1212	489900	571/169	64	4	2
GS0418	*Alchemilla glaucescens* Wallr.	Rosaceae	2903	803205	270/133	75	6	3
GS0419	*Luzula sylvatica* (Huds.) Gaudin	Juncaceae	985	325503	471/125	10	2	3
GS0420	*Plantago atrata Hoppe*	Plantaginaceae	3598	944357	272/127	71	5	4
GS0421	*Leucojum vernum* L.	Amaryllidaceae	13043	2506460	191/125	61	5	3
GS0422	*Tephroseris longifolia* (Jacq.) Griseb. et Schenk	Asteraceae	14609	3380832	236/129	77	7	3
GS0423	*Sorbus aucuparia* L.	Rosaceae	2997	918414	341/141	76	11	2
GS0424	*Clematis alpina* (L.) Mill.	Ranunculaceae	8299	1897137	235/125	74	8	4
GS0425	*Myosotis alpestris* F. W. Schmidt	Boraginaceae	2977	986144	401/145	23	26	3
GS0426	*Hieracium cymosum* L.	Asteraceae	8942	2213635	255/128	76	10	4
GS0427	*Gentiana verna* L.	Gentianaceae	3655	924772	253/128	75	4	3
GS0428	*Soldanella alpina* L. subsp. *alpina*	Primulaceae	8996	1883652	201/125	65	7	3
GS0431	*Galanthus nivalis* L.	Amaryllidaceae	12765	2503589	197/125	69	4	2
GS0432	*Melica uniflora* Retz.	Poaceae	2429	619761	255/125	76	4	4
GS0435	*Geranium pyrenaicum* Burm. fil. subsp. *pyrenaicum*	Geraniaceae	10440	1938928	172/125	55	4	3
GS0436	*Achillea millefolium* L.	Asteraceae	3404	934400	292/125	75	5	3
GS0437	*Rumex acetosella* L.	Polygonaceae	2231	658885	309/137	76	6	3
GS0438	*Viscaria vulgaris* Röhling	Caryophyllaceae	1828	602601	342/158	49	5	2
GS0439	*Ornithogalum umbellatum* L.	Hyacinthaceae	29052	5974178	206/125	74	6	3
GS0441	*Stellaria nemorum* L.	Caryophyllaceae	5843	1643078	306/132	58	21	1
GS0442	*Thalictrum aquilegiifolium* L.	Ranunculaceae	12335	3081866	247/125	78	33	3
GS0443	*Cirsium alsophilum* (Pollini) Soldano	Asteraceae	5837	1638558	302/131	79	20	3
GS0444	*Matteuccia struthiopteris* (L.) Tod.	Dryopteridaceae	6253	1920262	347/135	9	1	3
GS0445	*Chrysosplenium alternifolium* L.	Saxifragaceae	20768	6627207	346/184	71	34	2
GS0446	*Luzula nivea* (L.) DC.	Juncaceae	1895	544933	307/125	5	2	3
GS0447	*Atocion rupestre* (L.) Rafin.	Caryophyllaceae	3884	1082968	298/134	71	6	3
GS0448	*Selaginella helvetica* (L.) Spring	Selaginellaceae	1177	564145	1052/170	0	1	2
GS0449	*Veronica chamaedrys* L.	Plantaginaceae	1717	531459	324/135	67	6	2
GS0450	*Gymnocarpium dryopteris* (L.) Newman	Cystopteridaceae	3953	1324060	389/154	9	1	3
GS0452	*Lathyrus linifolius* (Reichard) Bässler	Fabaceae	8499	2098129	249/134	57	9	3
GS0453	*Knautia arvensis* (L.) Coult.	Dipsacaceae	15505	3315017	214/125	64	9	3
GS0454	*Avenella flexuosa* (L.) Parl.	Poaceae	4204	1203080	315/139	70	5	3
GS0455	*Rubus saxatilis* L.	Rosaceae	4058	1136722	287/143	73	9	3

GS0456	*Potentilla erecta* (L.) Raeusch.	Rosaceae	3859	1101755	301/135	71	7	3
GS0457	*Saxifraga paniculata* Mill.	Saxifragaceae	6367	1917486	314/137	52	30	3
GS0458	*Clematis recta* L.	Ranunculaceae	8644	1886277	217/125	76	5	3
GS0459	*Geranium pusillum* L.	Geraniaceae	6649	1720118	254/125	49	25	2
GS0460	*Geranium sanguineum* L.	Geraniaceae	7876	1952969	256/127	56	5	2
GS0462	*Briza media* L.	Poaceae	4662	1168529	258/126	71	5	3
GS0463	*Centaurea nigrescens* Willd.	Asteraceae	3686	1109304	313/155	78	6	3
GS0465	*Polygala nicaeensis* W. D. J. Koch	Polygalaceae	2162	768928	437/158	18	12	4
GS0466	*Cotinus coggygria* Scop.	Anacardiaceae	71191	17660238	238/189	70	36	0
GS0467	*Chaerophyllum temulum* L.	Apiaceae	10130	1976951	192/125	77	9	3
GS0468	*Carex baldensis* L.	Cyperaceae	693	321123	873/169	61	4	3
GS0469	*Stipa eriocaulis Borbás*	Poaceae	14069	3349010	241/127	77	12	2
GS0470	*Plantago argentea* Chaix	Plantaginaceae	8297	1960198	237/125	73	5	4
GS0471	*Euphorbia nicaeensis* All.	Euphorbiaceae	6543	1580654	242/125	46	6	4
GS0472	*Helleborus niger* L.	Ranunculaceae	5872	1314809	226/125	54	7	4
GS0473	*Inula ensifolia* L.	Asteraceae	6433	1957591	341/144	76	16	3
GS0474	*Inula hirta* L.	Asteraceae	9626	2553426	280/134	79	13	4
GS0475	*Euphorbia variabilis* Ces.	Euphorbiaceae	4611	1305927	287/145	28	5	3
GS0476	*Dorycnium herbaceum* Vill.	Fabaceae	9089	2502560	289/135	71	24	3
GS0477	*Buphthalmum salicifolium* L.	Asteraceae	26457	5233175	193/125	78	30	3
GS0478	*Inula ensifolia* L.	Asteraceae	15073	3945824	267/130	71	39	2
GS0479	*Globularia nudicaulis* L.	Plantaginaceae	11174	3290313	278/153	77	36	3
GS0480	*Fumana ericifolia* Wallr.	Cistaceae	24441	5087215	200/125	41	31	2
GS0481	*Lomelosia graminifolia* (L.) Greuter et Burdet subsp. *graminifolia*	Dipsacaceae	33701	6944084	202/125	60	35	2
GS0482	*Silene saxifraga* L.	Caryophyllaceae	8575	2204805	274/126	68	12	2
GS0483	*Knautia velutina* Briq.	Dipsacaceae	11888	2852718	245/127	57	22	4
GS0484	*Achnatherum calamagrostis* (L.) P. Beauv.	Poaceae	15580	3298730	201/125	71	32	3
GS0485	*Matthiola fruticulosa* (L.) Maire	Brassicaceae	3820	1084815	295/143	76	5	5
GS0486	*Carex mucronata All.*	Cyperaceae	1614	566713	401/160	69	4	3
GS0487	*Gentiana clusii* E. P. Perrier et Songeon	Gentianaceae	16917	3481419	206/125	78	7	3
GS0488	*Bellidiastrum michelii* Cass.	Asteraceae	11863	2768457	233/125	75	17	3
GS0489	*Sorbus aria* (L.) Crantz	Rosaceae	3899	1326913	430/141	75	33	3
GS0490	*Daphne reichsteinii* Landolt et Hauser	Thymelaeaceae	13053	3256087	256/130	49	21	3
GS0491	*Athamanta cretensis* L.	Apiaceae	6186	1416697	231/125	74	13	1
GS0492	*Carex sempervirens* Vill.	Cyperaceae	2948	829071	299/142	58	4	2
GS0493	*Pinus sylvestris* L.	Pinaceae	1530	489068	343/144	68	3	3
GS0494	*Pinus nigra* J. F. Arnold	Pinaceae	2004	585037	296/137	61	3	3
GS0495	*Cotoneaster tomentosus* (Aiton) Lindl.	Rosaceae	917	404009	586/189	75	6	1
GS0496	*Laserpitium*	Apiaceae	10128	2458086	250/125	80	13	3

	peucedanoides L.							
GS0497	*Sorbus aria* (L.) Crantz	Rosaceae	1023	365145	407/155	76	6	2
GS0498	*Ilex aquifolium* L.	Aquifoliaceae	5279	1838553	442/140	76	40	3
GS0499	*Betula pendula* Roth	Betulaceae	2149	723965	330/190	76	8	3
GS0500	*Horminum pyrenaicum* L.	Lamiaceae	5882	1794713	295/141	76	33	1
GS0501	*Centaurea rhaetica Moritzi*	Asteraceae	9129	2433369	278/140	80	7	5
GS0502	*Fagus sylvatica* L. subsp. *sylvatica*	Fagaceae	9111	2617856	272/138	76	34	2
GS0503	*Coronilla vaginalis* Lam.	Fabaceae	13455	3581854	277/126	71	38	3
GS0504	*Urtica dioica* L. subsp. *dioica*	Urticaceae	9069	1998409	221/125	51	13	2
GS0505	*Populus tremula* L.	Salicaceae	766	319643	650/151	76	6	3
GS0506	*Paederota bonarota* (L.) L.	Plantaginaceae	2877	1028239	473/155	72	23	4
GS0507	*Viola pinnata* L.	Violaceae	4801	1752120	455/151	74	38	3
GS0509	*Primula spectabilis* Tratt.	Primulaceae	11437	2787477	236/125	73	41	3
GS0510	*Valeriana saxatilis* L.	Valerianaceae	2386	1888370	3733/218	49	28	3
GS0511	*Gymnocarpium robertianum* (Hoffm.) Newman	Cystopteridaceae	6390	1988508	347/143	12	1	1
GS0512	*Acinos alpinus* (L.) Moench	Lamiaceae	4126	1200863	311/136	70	23	2
GS0513	*Daphne petraea* Leyb.	Thymelaeaceae	4947	1768652	440/148	49	36	4
GS0514	*Physoplexis comosa* (L.) Schur	Campanulaceae	4387	1259532	319/130	63	12	3
GS0515	*Genista radiata* (L.) Scop.	Fabaceae	13354	3943744	329/130	74	37	3
GS0516	*Rosa pendulina* L.	Rosaceae	5030	1338785	267/139	84	11	3
GS0517	*Kernera saxatilis* (L.) Sweet subsp. *saxatilis*	Brassicaceae	1485	656594	866/139	78	23	2
GS0518	*Linaria vulgaris* Mill. subsp. *vulgaris*	Plantaginaceae	5233	1577540	325/142	82	19	4
GS0519	*Orobanche gracilis* Sm.	Orobanchaceae	27322	7589525	285/125	19	29	1
GS0520	*Verbascum lychnitis* L.	Scrophulariaceae	7522	2280468	277/144	75	38	3
GS0521	*Carex austroalpina Becherer*	Cyperaceae	7289	1859088	256/134	59	4	2
GS0523	*Stachys alopecuros* (L.) Benth.	Lamiaceae	6642	1453683	218/125	82	7	2
GS0525	*Linum alpinum Jacq.*	Linaceae	21704	4539209	205/125	9	35	2
GS0526	*Asperula aristata* L. fil.	Rubiaceae	2848	773559	271/139	58	7	3
GS0527	*Phyteuma orbiculare* L.	Campanulaceae	9558	2270322	241/126	74	8	3
GS0528	*Erythronium dens-canis* L.	Liliaceae	2423	724947	301/143	62	5	1
GS0529	*Muscari botryoides* (L.) Mill.	Hyacinthaceae	30296	6511503	216/125	68	23	3
GS0530	*Veronica urticifolia* Jacq.	Plantaginaceae	6239	1682554	296/125	70	28	3
GS0531	*Laserpitium siler* L.	Apiaceae	7762	1933350	255/125	80	7	3
GS0532	*Dryopteris filix-mas* (L.) Schott	Dryopteridaceae	16936	4543733	288/126	13	2	2
GS0534	*Valeriana officinalis* L.	Valerianaceae	14063	2983087	214/125	47	6	2
GS0535	*Heracleum mantegazzianum* Sommier et Levier	Apiaceae	15398	3314569	214/125	78	6	4
GS0536	*Rosa canina* L.	Rosaceae	4435	1193592	270/140	78	13	4
GS0537	*Sagina apetala* Ard.	Caryophyllaceae	2754	858421	296/188	72	7	2

GS0539	*Senecio inaequidens* DC.	Asteraceae	15233	3737296	246/127	78	37	3
GS0541	*Lepidium ruderale* L.	Brassicaceae	1029	441948	665/147	78	10	4
GS0542	*Chaenorhinum minus* (L.) Lange	Plantaginaceae	1848	990674	1454/183	73	37	2
GS0544	*Dianthus sylvestris* Wulfen	Caryophyllaceae	6356	2116586	396/151	41	11	3
GS0545	*Galium lucidum* All.	Rubiaceae	4547	1203711	270/125	48	9	3
GS0546	*Festuca valesiaca* Schleich. ex Gaudin subsp. *valesiaca*	Poaceae	13107	3064455	238/125	78	3	3
GS0547	*Trifolium alpestre* L.	Fabaceae	8603	2478044	317/135	54	24	2
GS0548	*Poa nemoralis* L.	Poaceae	16465	3759519	230/125	74	7	3
GS0549	*Avenula praeusta* (Rchb.) Holub	Poaceae	12180	2556092	212/125	76	3	2
GS0550	*Festuca stricta Host*	Poaceae	6014	1538381	217/127	69	4	3
GS0552	*Koeleria cristata* (L.) Roem. et Schult.	Poaceae	25163	5403972	220/125	27	7	3
GS0553	*Convolvulus arvensis* L.	Convolvulaceae	5907	1760837	328/130	66	36	2
GS0554	*Ballota nigra* L.	Lamiaceae	1680	529968	343/136	71	5	2
GS0555	*Sempervivum arachnoideum* L.	Crassulaceae	9198	2690123	268/193	73	30	2
GS0556	*Hieracium pilosella* L.	Asteraceae	11793	2924238	254/127	77	19	2
GS0558	*Genista tinctoria* L.	Fabaceae	10858	2917419	278/125	74	35	1
GS0559	*Prunus spinosa* L. subsp. *spinosa*	Rosaceae	2262	676930	306/132	76	6	4
GS0560	*Asplenium septentrionale* (L.) Hoffm. subsp. *septentrionale*	Aspleniaceae	9989	2535279	267/125	7	3	1
GS0561	*Campanula glomerata* L.	Campanulaceae	5689	1445751	265/126	80	3	3
GS0562	*Trisetaria flavescens* (L.) Baumg.	Poaceae	10121	2566338	264/125	77	3	4
GS0563	*Juncus inflexus* L.	Juncaceae	3877	1490087	690/136	34	19	2
GS0564	*Melica transsilvanica Schur*	Poaceae	7700	1831279	243/125	76	5	4
GS0565	*Astragalus glycyphyllos* L.	Fabaceae	6958	2195941	377/130	74	31	3
GS0566	*Sanguisorba minor* Scop.	Rosaceae	2611	1008494	491/169	75	25	3
GS0568	*Galium boreale* L.	Rubiaceae	4509	1361253	302/165	19	20	1
GS0569	*Schedonorus pratensis* (Huds.) P. Beauv.	Poaceae	10955	2567533	239/125	77	3	4
GS0570	*Anthericum liliago* L.	Asparagaceae	25902	5345447	208/125	68	9	0
GS0571	*Luzula luzuloides* (Lam.) Dandy et Wilmott	Juncaceae	4746	1108083	228/125	5	2	3
GS0572	*Scirpus sylvaticus* L.	Cyperaceae	3372	1074949	301/197	76	6	2
GS0573	*Rubus hirtus (aggregato)*	Rosaceae	4408	1440517	410/134	66	37	4
GS0574	*Poa compressa* L.	Poaceae	25778	5726412	224/125	76	8	3
GS0575	*Dryopteris affinis* (Lowe) Fraser-Jenk. subsp. *affinis*	Dryopteridaceae	15551	4140660	287/128	12	1	3
GS0576	*Asplenium viride* Huds.	Aspleniaceae	6655	1925566	333/125	17	2	1
GS0577	*Lonicera alpigena* L. subsp. *alpigena*	Caprifoliaceae	10369	2715977	264/143	76	11	2
GS0579	*Lycopodium annotinum* L. subsp. *annotinum*	Lycopodiaceae	3057	894440	312/131	9	2	2
GS0580	*Huperzia selago* (L.) Bernh. ex Schrank et Mart. subsp. *selago*	Lycopodiaceae	3604	1197810	365/161	85	2	2

GS0581	*Lonicera nigra* L.	Caprifoliaceae	9509	3348125	428/148	64	35	2
GS0582	*Dryopteris dilatata* (Hoffm.) A. Gray	Dryopteridaceae	10754	2968174	298/126	23	1	4
GS0583	*Stellaria nemorum* L.	Caryophyllaceae	7795	1803643	232/125	43	15	3
GS0585	*Pyrola secunda* L.	Ericaceae	1187	434790	423/176	29	4	4
GS0586	*Rhamnus cathartica* L.	Rhamnaceae	19854	4992009	244/128	40	40	3
GS0587	*Geum rivale* L.	Rosaceae	2760	802589	321/126	73	12	3
GS0589	*Salvia verticillata* L. subsp. *verticillata*	Lamiaceae	5102	1624226	342/141	74	34	2
GS0591	*Acer pseudoplatanus* L.	Aceraceae	5724	2051231	537/133	78	31	3
GS0592	*Holandrea schottii* (Besser ex DC.) Reduron, Charpin et Pimenov	Apiaceae	9222	2180178	238/125	77	6	3
GS0593	*Festuca alpestris* Roem. et Schult.	Poaceae	10510	2603482	257/126	77	4	3
GS0594	*Tilia cordata* Mill.	Malvaceae	3316	1086360	332/191	79	8	3
GS0595	*Cephalanthera damasonium* (Mill.) Druce	Orchidaceae	11787	2810044	242/130	63	10	3
GS0596	*Moehringia bavarica* (L.) Gren.	Caryophyllaceae	8260	2175011	271/126	66	16	3
GS0598	*Knautia drymeia* Heuff.	Dipsacaceae	1440	460607	351/155	50	2	2
GS0599	*Silene vulgaris* (Moench) Garcke	Caryophyllaceae	13146	2963639	222/125	75	24	4
GS0600	*Lathyrus sylvestris* L. subsp. *sylvestris*	Fabaceae	7587	1749670	230/125	68	6	3
GS0601	*Galium aparine* L.	Rubiaceae	280	217129	5543/220	17	6	3
GS0602	*Rosa montana* Chaix	Rosaceae	4120	1105443	268/137	75	13	4
GS0603	*Campanula trachelium* L. subsp. *trachelium*	Campanulaceae	6850	1675392	249/125	78	4	4
GS0604	*Cirsium erisithales* (Jacq.) Scop.	Asteraceae	7893	2107553	277/125	77	30	3
GS0605	*Stellaria graminea* L.	Caryophyllaceae	3906	1218342	338/158	67	11	2
GS0606	*Prunella vulgaris* L. subsp. *vulgaris*	Lamiaceae	4579	1270144	275/150	74	8	3
GS0607	*Aegopodium podagraria* L.	Apiaceae	11073	2361048	213/125	77	7	3
GS0608	*Laburnum alpinum* (Mill.) Bercht. et J. Presl	Fabaceae	6616	2111105	385/144	75	17	2
GS0609	*Bromopsis benekenii* (Lange) Holub	Poaceae	7819	1798031	234/125	75	8	3
GS0610	*Lilium bulbiferum* L.	Liliaceae	1730	576008	373/148	76	5	2
GS0611	*Dactylorhiza maculata* (L.) Soó	Orchidaceae	14891	3932530	277/135	77	11	3
GS0612	*Ophioglossum vulgatum* L.	Ophioglossaceae	5011	1619819	380/153	82	6	0
GS0613	*Polygala vulgaris* L.	Polygalaceae	1841	593025	309/180	19	5	2
GS0614	*Linum catharticum* L.	Linaceae	4312	1406515	419/131	53	24	3
GS0615	*Euphrasia rostkoviana Hayne*	Orobanchaceae	7934	2315070	339/125	38	33	2
GS0617	*Traunsteinera globosa* (L.) Rchb.	Orchidaceae	7241	2107774	326/140	64	15	3
GS0618	*Cynosurus cristatus* L.	Poaceae	25429	4829061	188/125	74	5	3
GS0619	*Holcus lanatus* L.	Poaceae	26513	5136967	192/125	76	10	3
GS0620	*Paradisea liliastrum* (L.) Bertol.	Asparagaceae	14351	2937331	210/125	72	5	2
GS0621	*Orchis mascula* (L.) L.	Orchidaceae	21069	4516203	218/125	68	7	3
GS0622	*Cirsium acaule* Scop.	Asteraceae	28433	6288934	221/125	80	34	2

	subsp. *acaule*							
GS0623	*Rubus idaeus* L. subsp. *idaeus*	Rosaceae	7775	2055476	262/128	74	22	3
GS0624	*Bellardiochloa variegata* (Lam.) Kerguélen	Poaceae	25596	5166450	205/125	66	7	1
GS0625	*Rhinanthus freynii* (Sterneck) Fiori	Orobanchaceae	51250	9446006	179/125	32	32	2
GS0626	*Leontodon hispidus* L.	Asteraceae	2697	974478	436/164	77	16	4
GS0628	*Gymnadenia conopsea* (L.) R. Br.	Orchidaceae	10787	2531586	234/131	76	6	3
GS0629	*Platanthera bifolia* (L.) Rich.	Orchidaceae	6202	1507666	241/125	77	4	3
GS0630	*Gentianella germanica* (Willd.) E. F. Warb.	Gentianaceae	5874	1463230	259/129	69	13	3
GS0631	*Pyrola minor* L.	Ericaceae	220	190990	2115/283	28	5	3
GS0633	*Potentilla aurea* L. subsp. *aurea*	Rosaceae	3213	1260367	558/152	73	32	2
GS0634	*Noccaea praecox* (Wulfen) F. K. Meyer	Brassicaceae	1636	710513	738/154	79	30	3
GS0635	*Dianthus carthusianorum* L.	Caryophyllaceae	7058	1737028	242/137	39	6	3
GS0636	*Pimpinella saxifraga* L.	Apiaceae	9266	2074972	224/125	79	8	2
GS0637	*Centaurea scabiosa* L.	Asteraceae	9776	2649048	284/134	79	25	4
GS0639	*Phegopteris connectilis* (Michx.) Watt	Thelypteridaceae	3888	1401672	438/169	24	1	3
GS0640	*Polystichum braunii* (Spenn.) Fée	Dryopteridaceae	15846	4348717	296/132	73	1	3
GS0641	*Saxifraga rotundifolia* L. subsp. *rotundifolia*	Saxifragaceae	1141	428090	435/162	74	7	4
GS0642	*Prunus padus* L.	Rosaceae	1878	676974	343/200	78	9	4
GS0643	*Salix myrsinifolia* Salisb.	Salicaceae	482	251121	1571/172	73	6	3
GS0644	*Stellaria nemorum* L.	Caryophyllaceae	3548	1035999	314/137	56	22	3
GS0645	*Carduus personata* (L.) Jacq.	Asteraceae	4437	1129890	257/126	85	7	3
GS0646	*Lysimachia punctata* L.	Primulaceae	3654	1007304	276/144	80	9	4
GS0647	*Polygonatum verticillatum* (L.) All.	Asparagaceae	15323	3766812	253/130	76	8	1
GS0648	*Ribes petraeum* Wulfen	Grossulariaceae	5924	1944712	392/148	64	25	3
GS0649	*Filipendula ulmaria* (L.) Maxim.	Rosaceae	1356	442724	325/140	29	5	3
GS0650	*Fraxinus excelsior* L. subsp. *excelsior*	Oleaceae	2822	1312929	714/197	75	22	3
GS0651	*Onobrychis montana* DC. subsp. *montana*	Fabaceae	9159	2344122	260/130	67	8	3
GS0652	*Caltha palustris* L.	Ranunculaceae	16315	3166254	194/125	59	5	4
GS0653	*Carex spicata* Huds.	Cyperaceae	791	363235	644/189	73	4	1
GS0657	*Thesium alpinum* L.	Thesiaceae	6182	1897938	320/136	52	36	2
GS0660	*Scrophularia nodosa* L.	Scrophulariaceae	5646	1862984	385/168	70	15	4
GS0666	*Hippophae fluviatilis* (Soest) Rivas Mart.	Elaeagnaceae	11704	2540753	208/125	67	26	4
GS0667	*Saxifraga paniculata* Mill.	Saxifragaceae	2046	721826	391/169	57	8	3
GS0669	*Carduus defloratus* L.	Asteraceae	5940	1663460	301/134	85	14	2
GS0670	*Hieracium lachenalii* C. C. Gmel.	Asteraceae	13748	3200191	236/126	78	14	4
GS0671	*Picea abies* (L.) H. Karst.	Pinaceae	1552	530915	393/158	71	5	2
GS0672	*Hieracium murorum* L.	Asteraceae	12082	2985795	251/128	78	25	4
GS0673	*Polygala amara* L. subsp.	Polygalaceae	2985	908482	304/170	30	8	3

	brachyptera (Chodat) Hayek							
GS0674	*Digitalis grandiflora* Mill.	Plantaginaceae	4882	1678569	449/143	77	30	3
GS0675	*Digitalis lutea* L.	Plantaginaceae	5887	2029639	429/143	77	33	3
GS0676	*Trifolium aureum* Pollich subsp. *aureum*	Fabaceae	23728	5863900	230/156	75	32	4
GS0677	*Ranunculus montanus* Willd.	Ranunculaceae	15089	3408527	227/125	78	7	3
GS0678	*Viola biflora* L.	Violaceae	4286	1263268	314/146	74	8	3
GS0679	*Luzula luzulina* (Vill.) Dalla Torre et Sarnth.	Juncaceae	5782	1218155	203/125	7	2	3
GS0680	*Milium effusum* L.	Poaceae	3466	1005850	325/125	72	5	3
GS0681	*Phleum rhaeticum* (Humphries) Rauschert	Poaceae	27083	6207215	232/125	77	17	1
GS0682	*Pyrola uniflora* L.	Ericaceae	615	267740	550/213	31	2	3
GS0684	*Vaccinium myrtillus* L.	Ericaceae	1428	617549	704/154	75	15	3
GS0685	*Vaccinium vitis-idaea* L.	Ericaceae	1850	766957	467/209	58	13	3
GS0688	*Geum montanum* L.	Rosaceae	2208	683176	338/138	72	11	3
GS0689	*Carex ericetorum* Pollich	Cyperaceae	1603	535714	377/159	56	2	3
GS0690	*Nardus stricta* L.	Poaceae	9576	2386973	254/125	75	9	3
GS0692	*Lotus alpinus* (DC.) Schleicher	Fabaceae	8018	2314745	318/135	74	27	3
GS0693	*Pulsatilla vernalis* (L.) Mill. var. alpestris Aich. et Schw.	Ranunculaceae	7148	1711906	243/126	70	5	3
GS0694	*Primula elatior* (L.) Hill	Primulaceae	11692	2888950	236/125	75	38	3
GS0695	*Poa chaixii* Vill.	Poaceae	14256	3663112	273/125	75	26	3
GS0696	*Pulsatilla alpina* (L.) Delarbre	Ranunculaceae	6475	1864305	345/125	68	16	4
GS0697	*Ranunculus breyninus* Crantz	Ranunculaceae	10714	2372371	219/125	79	8	3
GS0698	*Valeriana montana* L.	Valerianaceae	19791	5190272	249/125	56	29	3
GS0699	*Pedicularis verticillata* L.	Orobanchaceae	6902	1519954	221/125	57	6	2
GS0700	*Bistorta vivipara* (L.) Delarbre	Polygonaceae	14737	3368208	228/125	73	26	2
GS0702	*Trifolium hybridum* L.	Fabaceae	26583	6356029	234/128	71	33	2
GS0703	*Polypodium vulgare* L.	Polypodiaceae	3535	1274809	419/171	11	1	3
GS0704	*Prenanthes purpurea* L.	Asteraceae	9561	2280098	242/125	78	6	3
GS0705	*Arctium minus* (Hill) Bernh.	Asteraceae	6655	1762063	276/133	85	9	3
GS0706	*Deschampsia cespitosa* (L.) P. Beauv.	Poaceae	8575	2074254	250/125	75	4	3
GS0707	*Listera ovata* (L.) R. Br.	Orchidaceae	4440	1105173	252/125	77	5	3
GS0708	*Viburnum opulus* L.	Adoxaceae	11211	2561061	234/125	60	11	4
GS0709	*Sedum rupestre* L.	Crassulaceae	13424	3319000	251/132	75	9	3
GS0710	*Cerastium arvense* L.	Caryophyllaceae	7540	1967727	286/125	55	28	2
GS0712	*Gentiana cruciata* L. subsp. *cruciata*	Gentianaceae	11419	3133126	288/136	76	30	3
GS0713	*Veratrum album* L.	Melanthiaceae	72882	14582032	200/125	76	38	1
GS0717	*Cirsium erisithales* (Jacq.) Scop.	Asteraceae	8742	2330976	269/128	86	34	3
GS0718	*Chaerophyllum hirsutum* L.	Apiaceae	20607	4152676	204/125	70	9	4
GS0719	*Crepis paludosa* (L.) Moench	Asteraceae	12448	2797255	226/125	78	7	4
GS0720	*Cypripedium calceolus* L.	Orchidaceae	910	391648	590/188	74	5	3

GS0722	*Ranunculus platanifolius* L.	Ranunculaceae	31879	6417313	203/125	77	7	4
GS0723	*Streptopus amplexifolius* (L.) DC.	Asparagaceae	24149	4993916	205/125	70	19	0
GS0724	*Scorzonera aristata* Ramond ex DC.	Asteraceae	9764	2292506	238/125	76	6	3
GS0725	*Astrantia major* L.	Apiaceae	10873	2556582	232/125	75	21	1
GS0727	*Laserpitium krapfii Crantz* subsp. *gaudinii* (Moretti) Thell.	Apiaceae	1598	584058	439/158	81	6	3
GS0728	*Gentiana lutea* L.	Gentianaceae	14943	3582268	241/125	71	33	3
GS0730	*Eriophorum latifolium Hoppe*	Cyperaceae	6284	1651127	258/145	66	4	2
GS0731	*Primula farinosa* L.	Primulaceae	11857	3124479	275/130	66	28	4
GS0732	*Menyanthes trifoliata* L.	Menyanthaceae	4555	1653516	536/134	52	35	1
GS0733	*Bartsia alpina* L.	Orobanchaceae	7292	2161312	329/125	40	39	4
GS0734	*Pinguicula vulgaris* L.	Lentibulariaceae	4419	1377844	403/125	57	36	3
GS0735	*Carex rostrata Stokes*	Cyperaceae	1341	403827	328/156	34	2	2
GS0736	*Myrrhis odorata* (L.) Scop.	Apiaceae	7340	1592277	217/125	72	10	3
GS0737	*Hypochaeris uniflora* Vill.	Asteraceae	8057	2010813	258/130	77	6	3
GS0738	*Trifolium alpinum* L.	Fabaceae	6992	1960175	320/126	66	11	3
GS0739	*Arnica montana* L. subsp. *montana*	Asteraceae	10448	2931670	296/134	77	30	3
GS0740	*Pulsatilla alpina* (L.) Delarbre	Ranunculaceae	542	302577	1062/198	70	5	4
GS0741	*Campanula barbata* L.	Campanulaceae	2851	802412	299/133	79	3	4
GS0742	*Sorbus chamaemespilus* (L.) Crantz	Rosaceae	6097	1779948	317/128	73	34	3
GS0743	*Silene nemoralis* Waldst. et Kit.	Caryophyllaceae	5001	1222962	249/125	68	7	3
GS0744	*Orlaya grandiflora* (L.) Hoffm.	Apiaceae	17710	3974185	225/125	79	30	4
GS0745	*Petrorhagia saxifraga* (L.) Link	Caryophyllaceae	5134	1508128	315/131	36	28	2
GS0746	*Lolium multiflorum* Lam.	Poaceae	5665	1423963	258/125	77	3	3
GS0748	*Sedum sexangulare* L.	Crassulaceae	487	289327	1193/209	80	7	4
GS0749	*Euphorbia maculata* L.	Euphorbiaceae	8153	2225133	297/125	31	19	3
GS0750	*Calendula officinalis* L.	Asteraceae	6478	1818376	292/125	78	36	3
GS0751	*Sagina procumbens* L.	Caryophyllaceae	3113	850880	250/154	75	6	3
GS0752	*Polygonum aviculare* L.	Polygonaceae	19840	4440746	217/125	75	36	3
GS0753	*Sedum album* L.	Crassulaceae	2615	830093	329/152	72	11	2
GS0754	*Antirrhinum majus* L.	Plantaginaceae	1552	997298	1787/204	67	36	3
GS0755	*Verbena officinalis* L.	Verbenaceae	7539	2090828	262/125	76	40	2
GS0756	*Satureja hortensis* L.	Lamiaceae	5925	1680908	313/125	67	26	3
GS0757	*Hypericum perforatum* L.	Hypericaceae	2985	987846	364/173	12	11	3
GS0758	*Tanacetum vulgare* L.	Asteraceae	14064	3383433	247/125	77	21	3
GS0760	*Clinopodium vulgare* L.	Lamiaceae	1968	685265	402/159	73	12	3
GS0761	*Oreoselinum nigrum Delarbre*	Apiaceae	10908	2674491	248/125	77	36	3
GS0762	*Melampyrum pratense* L.	Orobanchaceae	8294	2111468	269/125	32	18	4
GS0763	*Galium pumilum Murray*	Rubiaceae	4948	1383173	283/149	44	11	3
GS0764	*Tanacetum corymbosum* (L.) Sch. Bip.	Asteraceae	9663	2423952	261/130	72	11	3
GS0765	*Anacamptis pyramidalis*	Orchidaceae	2366	738531	329/151	77	4	3

	(L.) Rich.							
GS0766	*Lathyrus pratensis* L.	Fabaceae	9328	2019500	218/125	48	5	2
GS0767	*Festuca heterophylla* Lam. subsp. *heterophylla*	Poaceae	4074	1122824	292/129	75	4	3
GS0768	*Lactuca muralis* (L.) Gaertn.	Asteraceae	4773	1524128	380/130	77	31	3
GS0769	*Campanula persicifolia* L.	Campanulaceae	9440	1928233	203/125	62	3	4
GS0770	*Phyteuma betonicifolium* Vill.	Campanulaceae	4950	1261714	266/125	54	4	3
GS0771	*Campanula rapunculoides* L. subsp. *rapunculoides*	Campanulaceae	3708	1053617	303/135	75	3	3
GS0772	*Epipactis atrorubens* (Hoffm.) Besser	Orchidaceae	3379	957552	299/137	76	5	3
GS0773	*Linum tenuifolium* L.	Linaceae	6142	1777140	342/125	53	14	4
GS0774	*Silene otites* (L.) Wibel	Caryophyllaceae	12602	2735361	218/125	63	4	3
GS0776	*Teucrium montanum* L.	Lamiaceae	16499	4148165	245/125	69	23	2
GS0777	*Teucrium chamaedrys* L.	Lamiaceae	3785	904837	231/125	71	6	3
GS0778	*Melampyrum cristatum* L. subsp. *cristatum*	Orobanchaceae	11207	2460864	220/125	35	12	1
GS0779	*Rosa arvensis* Huds.	Rosaceae	3293	989232	299/162	75	15	3
GS0782	*Carduus nutans* L.	Asteraceae	7977	2189537	283/125	79	33	4
GS0783	*Securigera varia* (L.) Lassen	Fabaceae	28581	6261510	216/125	72	35	2
GS0784	*Campanula rapunculus* L.	Campanulaceae	7333	2045081	310/129	52	19	3
GS0785	*Asparagus officinalis* L.	Asparagaceae	12277	3112686	253/125	73	33	3
GS0786	*Medicago falcata* L. subsp. *falcata*	Fabaceae	13164	3207364	243/125	72	25	4
GS0787	*Astragalus onobrychis* L.	Fabaceae	6486	1891734	317/149	72	21	3
GS0788	*Potentilla recta* L.	Rosaceae	2420	831612	380/171	74	15	3
GS0789	*Centaurea scabiosa* L.	Asteraceae	7197	1967048	287/139	74	19	3
GS0790	*Petrorhagia prolifera* (L.) P. W. Ball et Heywood	Caryophyllaceae	13565	3170648	229/125	49	26	3
GS0791	*Castanea sativa* Mill.	Fagaceae	4035	1194598	313/140	78	17	3
GS0792	*Potentilla alba* L.	Rosaceae	1701	564793	386/139	78	6	3
GS0793	*Fragaria moschata Duchesne*	Rosaceae	12676	3446974	244/189	77	37	3
GS0794	*Galium aparine* L.	Rubiaceae	977	504408	890/203	52	14	3
GS0795	*Frangula alnus* Mill. subsp. *alnus*	Rhamnaceae	53834	12957125	233/163	62	40	3
GS0796	*Veronica officinalis* L.	Plantaginaceae	3483	973793	299/129	28	6	4
GS0797	*Hypericum montanum* L.	Hypericaceae	5830	1772066	311/141	12	22	3
GS0798	*Lychnis flos-cuculi* L.	Caryophyllaceae	17861	3474898	195/125	70	6	3
GS0799	*Hieracium lactucella* Wallr.	Asteraceae	15134	3618574	245/125	77	12	4
GS0800	*Carex pallescens* L.	Cyperaceae	519	260875	886/190	60	3	3
GS0802	*Phleum pratense* L.	Poaceae	11240	2653483	244/125	81	3	4
GS0803	*Sherardia arvensis* L.	Rubiaceae	1232	489906	526/172	26	8	3
GS0804	*Anagallis arvensis* L.	Primulaceae	1872	1077947	1982/189	70	32	3
GS0805	*Galium verum* L.	Rubiaceae	1413	519507	428/164	18	5	3
GS0806	*Hemerocallis fulva* L.	Xanthorrhoeaceae	12348	2765923	227/125	71	8	2
GS0807	*Calystegia sepium* (L.) R. Br. subsp. *sepium*	Convolvulaceae	22321	4859369	213/125	65	37	3

GS0808	*Verbascum chaixii* Vill.	Scrophulariaceae	1600	936009	1282/207	72	29	3
GS0809	*Erigeron acris* L.	Asteraceae	9141	2362043	258/128	77	30	3
GS0810	*Melampyrum sylvaticum* L. subsp. *sylvaticum*	Orobanchaceae	16640	3576260	216/125	27	19	1
GS0811	*Melilotus officinalis* (L.) Pall.	Fabaceae	5753	1532030	285/125	55	19	3
GS0812	*Rhinanthus minor* L.	Orobanchaceae	12595	2979848	230/125	48	29	2
GS0813	*Anchusa officinalis* L.	Boraginaceae	3032	866178	288/134	48	15	3
GS0814	*Sisyrinchium montanum* Greene	Iridaceae	6515	1441097	220/125	77	6	1
GS0815	*Polygala amarella* Crantz	Polygalaceae	512	268347	737/214	22	3	3
GS0817	*Tofieldia calyculata* (L.) Wahlenb.	Tofieldiaceae	6822	2001133	342/128	35	20	3
GS0818	*Scorzonera humilis* L.	Asteraceae	7946	1940423	244/133	84	8	3
GS0819	*Alchemilla monticola Opiz*	Rosaceae	2456	703548	286/137	78	5	3
GS0820	*Pimpinella major* (L.) Huds.	Apiaceae	1923	562817	313/132	76	6	3
GS0821	*Tetragonolobus maritimus* (L.) Roth	Fabaceae	24292	5156592	209/125	74	33	2
GS0822	*Veronica beccabunga* L.	Plantaginaceae	5473	1437764	277/125	67	20	3
GS0823	*Cirsium palustre* (L.) Scop.	Asteraceae	6037	1569743	269/127	79	15	4
GS0824	*Carex paniculata* L.	Cyperaceae	602	314033	1088/174	65	4	2
GS0825	*Equisetum fluviatile* L.	Equisetaceae	10243	2337262	229/125	83	13	3
GS0826	*Tolpis staticifolia* (All.) Sch. Bip.	Asteraceae	4248	1355052	388/125	81	33	2
GS0827	*Rosa agrestis Savi*	Rosaceae	901	290273	343/167	52	2	4
GS0829	*Astragalus hypoglottis* L.	Fabaceae	3471	1139959	376/154	74	7	3
GS0830	*Hippocrepis comosa L.* subsp. *comosa*	Fabaceae	8447	2025514	247/125	68	7	3
GS0831	*Chenopodium album* L.	Amaranthaceae	7657	1752423	227/125	14	6	3
GS0833	*Orchis militaris* L.	Orchidaceae	5246	1294725	245/130	75	5	3
GS0834	*Campanula cespitosa* Scop.	Campanulaceae	7302	2023437	297/135	68	18	3
GS0837	*Neotinea ustulata* (L.) R. M. Bateman, Pridgeon et M. W. Chase	Orchidaceae	12496	3449628	286/151	69	14	2
GS0838	*Onobrychis arenaria* (Kit.) DC.	Fabaceae	13322	3423020	262/129	50	15	4
GS0839	*Ophrys insectifera* L.	Orchidaceae	6839	1711756	257/125	72	4	3
GS0840	*Equisetum ramosissimum* Desf.	Equisetaceae	2087	570285	280/125	82	1	5
GS0841	*Eleocharis quinqueflora* (Hartmann) O. Schwarz	Cyperaceae	1422	602292	559/191	38	4	3
GS0842	*Eleocharis uniglumis* (Link) Schult.	Cyperaceae	1491	498993	425/143	35	1	0
GS0843	*Carex lepidocarpa Tausch*	Cyperaceae	4708	1506873	337/191	63	9	2
GS0844	*Juncus compressus* Jacq.	Juncaceae	2778	796880	283/152	12	3	2
GS0845	*Geranium pratense* L. subsp. *pratense*	Geraniaceae	14381	3017130	205/125	55	7	1
GS0846	*Chenopodium hybridum* L.	Amaranthaceae	5491	1700341	298/147	69	33	3
GS0847	*Pedicularis hacquetii Graf*	Orobanchaceae	7473	2145472	316/125	73	37	2
GS0848	*Inula salicina* L.	Asteraceae	9619	2565972	274/134	78	26	3
GS0849	*Phyteuma ovatum Honck.*	Campanulaceae	4534	1162858	267/125	72	3	4

GS0850	*Pyrola secunda* L.	Ericaceae	1215	388694	347/167	30	4	4
GS0851	*Carex pilulifera* L.	Cyperaceae	1028	406731	492/176	69	3	2
GS0852	*Antennaria dioica* (L.) Gaertn.	Asteraceae	3682	1087902	319/143	77	4	4
GS0854	*Listera cordata* (L.) R. Br.	Orchidaceae	10092	2021957	198/125	67	5	3
GS0855	*Pinus cembra* L.	Pinaceae	4220	1129997	275/131	77	3	3
GS0856	*Carex leporina* L.	Cyperaceae	639	296312	868/181	71	3	4
GS0857	*Carex viridula Michx.*	Cyperaceae	2454	759484	298/188	59	4	2
GS0860	*Rumex alpinus* L.	Polygonaceae	11125	3092808	268/139	76	38	2
GS0861	*Alchemilla crinita Buser*	Rosaceae	1494	495045	371/137	74	5	3
GS0862	*Alchemilla glabra Neygenf.*	Rosaceae	2535	710504	282/129	71	7	3
GS0863	*Equisetum variegatum* Schleich. ex Weber et D. Mohr	Equisetaceae	14664	3103506	206/125	81	15	4
GS0864	*Salix mielichhoferi Saut.*	Salicaceae	143	156474	3451/372	68	6	2
GS0865	*Triglochin palustre* L.	Juncaginaceae	4673	1247558	276/125	62	9	3
GS0866	*Trifolium badium Schreb.*	Fabaceae	21363	4881071	224/125	73	34	2
GS0868	*Astragalus alpinus* L.	Fabaceae	13428	3342857	256/130	72	20	4
GS0869	*Glyceria notata Chevall.*	Poaceae	5248	1161050	223/125	76	5	2
GS0870	*Oxytropis jacquinii Bunge*	Fabaceae	12423	3344008	288/134	70	21	3
GS0871	*Rorippa palustris* (L.) Besser	Brassicaceae	2195	753332	362/188	76	12	2
GS0872	*Pedicularis elongata* A. Kern.	Orobanchaceae	7087	1988142	293/125	77	32	3
GS0873	*Astragalus australis* (L.) Lam.	Fabaceae	15747	4121075	279/131	73	25	3
GS0874	*Cirsium heterophyllum* (L.) Hill	Asteraceae	7454	1838375	247/131	60	10	4
GS0876	*Salix hastata* L.	Salicaceae	2730	862302	324/188	71	11	4
GS0877	*Salix waldsteiniana* Willd.	Salicaceae	379	178843	1406/125	62	5	1
GS0879	*Hedysarum hedysaroides* (L.) Schinz et Thell.	Fabaceae	11424	2779343	248/128	51	12	3
GS0880	*Oxytropis campestris* (L.) DC.	Fabaceae	12867	3166088	249/136	69	15	3
GS0881	*Leontodon helveticus* Mérat em. Widder	Asteraceae	17498	4370351	261/129	74	11	2
GS0883	*Barbarea bracteosa* Guss.	Brassicaceae	1410	468913	419/127	78	6	2
GS0884	*Veronica fruticans* Jacq.	Plantaginaceae	4291	1359199	376/130	81	31	3
GS0885	*Galium anisophyllon* Vill.	Rubiaceae	2983	898791	297/165	57	9	3
GS0886	*Thymus praecox Opiz*	Lamiaceae	2137	724429	458/133	65	13	2
GS0887	*Potentilla crantzii* (Crantz) Beck ex Fritsch subsp. *crantzii*	Rosaceae	2963	881170	307/150	63	9	3
GS0888	*Anthyllis vulneraria* L.	Fabaceae	13537	3306959	239/125	71	35	4
GS0889	*Lonicera caerulea* L. subsp. *caerulea*	Caprifoliaceae	10306	2690214	268/136	74	12	2
GS0890	*Carex ferruginea* Scop.	Cyperaceae	581	278350	1041/173	65	4	2
GS0891	*Imperatoria ostruthium* L.	Apiaceae	11131	2696868	250/125	79	22	3
GS0893	*Festuca norica* (Hack.) K. Richt.	Poaceae	8048	1966961	255/125	76	4	4
GS0894	*Nigritella nigra* (L.)	Orchidaceae	13774	3132905	229/126	76	5	2

	Rchb. fil.							
GS0895	*Helianthemum nummularium* (L.) Mill.	Cistaceae	12031	2545800	213/125	26	8	2
GS0896	*Anthoxanthum nipponicum Honda*	Poaceae	5676	1437807	263/131	49	1	4
GS0897	*Festuca paniculata* (L.) Schinz et Thell. subsp. *paniculata*	Poaceae	11767	2847904	249/125	75	6	4
GS0898	*Hieracium hoppeanum Schult.*	Asteraceae	9551	2395105	258/130	78	11	4
GS0899	*Juncus triglumis* L.	Juncaceae	1799	464375	267/125	13	3	2
GS0900	*Eriophorum angustifolium Honck.*	Cyperaceae	942	425699	749/180	56	4	2
GS0901	*Eriophorum vaginatum* L.	Cyperaceae	4802	1268850	261/132	65	3	3
GS0902	*Carex hostiana* DC.	Cyperaceae	733	343474	792/181	66	4	2
GS0903	*Dactylorhiza majalis* (Rchb.) P. F. Hunt et Summerh.	Orchidaceae	17656	4353262	251/134	76	9	4
GS0904	*Pedicularis palustris* L. subsp. *palustris*	Orobanchaceae	6439	1800209	306/125	59	26	3
GS0905	*Allium schoenoprasum* L.	Amaryllidaceae	7780	2237127	314/138	69	14	3
GS0906	*Senecio abrotanifolius* L.	Asteraceae	12567	3157400	258/128	77	21	3
GS0907	*Salix glabra* Scop.	Salicaceae	1187	414579	345/173	76	6	4
GS0909	*Leucopoa pulchella* (Schrad.) H. Scholz et Foggi	Poaceae	7328	1880887	271/127	68	3	3
GS0910	*Juniperus communis* L.	Cupressaceae	888	400255	641/188	74	2	3
GS0912	*Pinguicula leptoceras Rchb.*	Lentibulariaceae	3758	1433433	953/125	52	33	3
GS0913	*Pedicularis tuberosa* L.	Orobanchaceae	1648	553721	332/180	81	9	2
GS0914	*Avenula versicolor* (Vill.) M. LaÃnz	Poaceae	4759	1210362	268/126	65	3	4
GS0915	*Festuca pseudovaria Vetter*	Poaceae	9059	2292315	264/132	66	4	4
GS0916	*Allium victorialis* L.	Amaryllidaceae	16570	3803458	232/125	71	12	3
GS0917	*Trifolium pratense* L.	Fabaceae	9825	2700316	274/137	47	30	2
GS0918	*Pseudorchis albida* (L.) Á. Löve et D. Löve	Orchidaceae	9265	3111307	463/128	68	29	3
GS0920	*Botrychium lunaria* (L.) Sw.	Ophioglossaceae	8945	1939403	211/125	5	1	1
GS0921	*Salix breviserrata* Flod.	Salicaceae	2190	861464	457/200	70	13	4
GS0922	*Gentiana punctata* L.	Gentianaceae	5606	1653786	326/133	74	21	3
GS0924	*Daphne striata* Tratt.	Thymelaeaceae	16506	3957163	239/125	52	33	2
GS0925	*Arctostaphylos alpinus* (L.) Spreng.	Ericaceae	60	144720	15588/1192	49	5	4
GS0926	*Sesleriella sphaerocephala* (Ard.) Deyl	Poaceae	6006	1548533	281/125	66	5	3
GS0927	*Cirsium spinosissimum* (L.) Scop.	Asteraceae	18431	4206045	224/125	77	38	3
GS0928	*Verbascum thapsus* L.	Scrophulariaceae	1407	832561	1298/209	76	30	4
GS0929	*Minuartia laricifolia* (L.) Schinz et Thell.	Caryophyllaceae	10227	2582112	269/134	50	7	3
GS0930	*Campanula spicata* L.	Campanulaceae	5859	1447521	254/125	78	4	4
GS0931	*Reseda lutea* L. subsp. *lutea*	Resedaceae	1934	645039	365/150	52	5	2
GS0933	*Minuartia rostrata* (Pers.)	Caryophyllaceae	6260	1623984	262/133	68	8	3

	Rchb.							
GS0934	*Aconitum lycoctonum* L. em. Koelle	Ranunculaceae	9349	2146547	229/129	77	5	3
GS0935	*Aconitum napellus* L. em. Skalický	Ranunculaceae	19036	4078138	217/125	75	6	3
GS0936	*Anisantha tectorum* (L.) Nevski	Poaceae	11670	2443938	205/125	76	5	3
GS0939	*Centaurea maculosa* Lam.	Asteraceae	11816	2823523	240/126	79	16	3
GS0940	*Linaria angustissima* (Loisel.) Borbás	Plantaginaceae	2888	884172	310/162	70	11	3
GS0941	*Solidago virgaurea* L.	Asteraceae	69165	14998796	218/125	78	31	2
GS0942	*Oxytropis halleri* W. D. J. Koch	Fabaceae	15324	3956515	269/138	73	18	3
GS0943	*Poa molinerii Balb.*	Poaceae	11020	4693998	583/202	49	1	3
GS0944	*Thalictrum foetidum* L.	Ranunculaceae	54991	12443906	223/126	71	37	2
GS0945	*Acinos arvensis* (Lam.) Dandy	Lamiaceae	7234	1859493	257/128	67	28	3
GS0947	*Chenopodium foliosum Asch.* subsp. *foliosum*	Chenopodiaceae	495	285792	1427/209	76	5	3
GS0949	*Carex supina Wahlenb.*	Cyperaceae	1979	628590	348/146	61	4	2
GS0950	*Achillea tomentosa* L.	Asteraceae	8004	1954520	253/125	80	9	3
GS0951	*Hyoscyamus niger* L.	Solanaceae	6126	1552783	259/129	79	8	3
GS0952	*Torilis japonica* (Houtt.) DC.	Apiaceae	15953	3835856	242/125	72	18	4
GS0953	*Arabis nova* Vill. subsp. *nova*	Brassicaceae	2269	711874	319/154	75	6	2
GS0954	*Onopordum acanthium* L. subsp. *acanthium*	Asteraceae	9098	2095529	228/125	80	17	4
GS0955	*Helianthemum nummularium* (L.) Mill.	Cistaceae	11642	2393562	209/125	14	9	1
GS0956	*Chenopodium glaucum* L.	Chenopodiaceae	16557	3914238	221/136	75	33	2
GS0957	*Reseda luteola* L.	Resedaceae	17060	3842362	215/125	52	25	3
GS0958	*Trifolium arvense* L.	Fabaceae	11379	2915252	255/131	65	30	2
GS0960	*Onosma helvetica* Boiss. em. Teppner	Boraginaceae	6364	1483980	239/125	46	9	3
GS0961	*Oxytropis pilosa* (L.) DC.	Fabaceae	12040	3143330	281/133	74	17	3
GS0962	*Campanula bononiensis* L.	Campanulaceae	9311	2402681	269/125	78	28	3
GS0963	*Melilotus albus Medik.*	Fabaceae	16427	3634537	213/125	64	37	3
GS0964	*Colutea arborescens* L.	Fabaceae	3704	1430674	528/162	68	37	3
GS0965	*Scabiosa triandra* L.	Dipsacaceae	7253	2020393	304/129	56	31	2
GS0966	*Dryopteris remota* (A. Braun ex Döll) Druce	Dryopteridaceae	16635	4285998	271/128	27	1	3
GS0967	*Circaea alpina* L. subsp. *alpina*	Onagraceae	1817	880840	859/188	62	19	3
GS0968	*Senecio alpinus* (L.) Scop.	Asteraceae	71344	28646644	425/208	80	6	2
GS0969	*Dryopteris affinis* (Lowe) Fraser-Jenk. subsp. *affinis*	Dryopteridaceae	9628	2828651	321/137	22	1	3
GS0970	*Juncus trifidus* L.	Juncaceae	4451	1081014	242/125	80	5	0
GS0971	*Festuca rubra* L.	Poaceae	16109	4514472	284/192	42	2	3
GS0972	*Knautia longifolia* (Waldst. et Kit.) W. D. J. Koch	Dipsacaceae	22210	6750471	314/193	61	6	2
GS0974	*Noccaea caerulescens* (J. et C. Presl.) F. K. Meyer	Brassicaceae	1020	317589	385/126	65	6	3

GS0975	*Chaerophyllum hirsutum* L.	Apiaceae	10966	2289956	209/125	75	6	3
GS0976	*Sedum annuum* L.	Crassulaceae	1370	422795	326/128	75	7	4
GS0977	*Leontodon autumnalis* L.	Asteraceae	9278	2388551	270/128	76	16	2
GS0978	*Epilobium collinum* C. C. Gmel.	Onagraceae	3065	959306	375/133	51	8	3
GS0979	*Hypericum maculatum* Crantz subsp. *maculatum*	Hypericaceae	29094	7289421	233/171	27	31	3
GS0980	*Bupleurum stellatum* L.	Apiaceae	18393	4196918	228/125	73	35	4
GS0982	*Galeopsis tetrahit* L.	Lamiaceae	7481	2078604	300/126	67	24	2
GS0983	*Alchemilla transiens* (Buser) Buser	Rosaceae	422	249752	2525/172	79	6	4
GS0984	*Aconitum tauricum Wulfen*	Ranunculaceae	18622	4018577	218/125	48	4	4
GS0985	*Saxifraga stellaris* L. subsp. *engleri* P. Fourn.	Saxifragaceae	4761	1623615	457/127	48	44	2
GS0986	*Carex frigida* All.	Cyperaceae	965	396458	664/159	61	4	2
GS0987	*Juncus articulatus* L.	Juncaceae	1739	505391	300/136	7	3	2
GS0988	*Juncus filiformis* L.	Juncaceae	5124	1471718	305/141	9	4	2
GS0989	*Cardamine pratensis* L.	Brassicaceae	14798	3484623	232/125	75	34	2
GS0990	*Plantago maritima* L.	Plantaginaceae	4334	1075630	250/125	80	5	3
GS0991	*Vaccinium uliginosum* L.	Ericaceae	3673	1302488	379/197	65	10	3
GS0993	*Agrostis alpina* Scop.	Poaceae	19640	4313867	223/125	78	7	3
GS0994	*Phyteuma hemisphaericum* L.	Campanulaceae	10423	2795286	285/126	70	36	3
GS0995	*Luzula lutea* (All.) DC.	Juncaceae	4728	1213838	251/138	4	3	4
GS0996	*Hieracium piliferum Hoppe*	Asteraceae	17038	3967905	239/125	77	11	4
GS0997	*Festuca halleri* All. subsp. *halleri*	Poaceae	5024	1414382	302/133	75	3	4
GS0998	*Hieracium alpinum* L.	Asteraceae	11773	2852557	250/130	78	7	4
GS0999	*Veronica bellidioides* L.	Plantaginaceae	1855	589664	331/143	35	6	3
GS1000	*Senecio incanus* L.	Asteraceae	14888	3698274	256/130	78	27	3
GS1001	*Silene acaulis* (L.) Jacq.	Caryophyllaceae	10861	2722136	256/133	70	6	4
GS1002	*Saxifraga exarata* Vill.	Saxifragaceae	3955	1627268	578/174	48	24	3
GS1004	*Festuca melanopsis Foggi,* Graz. Rossi et Signorini	Poaceae	7783	1908880	251/125	78	3	3
GS1006	*Erigeron uniflorus* L.	Asteraceae	8971	2225727	256/130	77	7	3
GS1007	*Leucanthemopsis alpina* (L.) Heywood	Asteraceae	15463	3266995	209/125	81	6	3
GS1008	*Antennaria carpathica* (Wahlenb.) Bluff et Fingerh.	Asteraceae	14856	3436093	236/125	79	6	4
GS1009	*Alchemilla flabellata Buser*	Rosaceae	347	223052	2502/194	81	6	3
GS1010	*Luzula alpino-pilosa* (Chaix) Breistr.	Juncaceae	1512	507052	391/152	8	2	3
GS1011	*Juncus jacquinii* L.	Juncaceae	19128	5158116	247/188	24	25	3
GS1012	*Minuartia rupestris* (Scop.) Schinz et Thell.	Caryophyllaceae	3861	1467948	487/160	69	29	3
GS1013	*Hieracium sphaerocephalum* Froel.	Asteraceae	2538	748210	298/144	75	6	4
GS1015	*Androsace vandellii* (Turra) Chiov.	Primulaceae	1001	723464	4205/196	59	34	2
GS1016	*Minuartia laricifolia* (L.) Schinz et Thell.	Caryophyllaceae	11551	3046375	282/137	37	19	4

GS1017	*Potentilla grandiflora* L.	Rosaceae	913	356323	622/153	68	4	4
GS1018	*Orchis mascula* (L.) L.	Orchidaceae	14090	3131279	224/125	63	6	3
GS1019	*Rhodiola rosea* L.	Crassulaceae	5418	1601128	327/136	64	24	3
GS1020	*Verbascum alpinum Turra*	Scrophulariaceae	998	752008	1886/254	75	29	3
GS1021	*Allium senescens* L. subsp. *montanum* (Fries) Holub	Amaryllidaceae	15405	3230760	214/125	71	5	3
GS1022	*Hieracium atratum* Fr.	Asteraceae	6616	1616762	244/130	78	8	4
GS1023	*Epilobium alsinifolium* Vill.	Onagraceae	3891	1144810	351/125	70	9	3
GS1024	*Gnaphalium sylvaticum* L.	Asteraceae	8414	1961169	231/125	83	4	4
GS1026	*Juncus effusus* L.	Juncaceae	1457	505071	364/188	12	3	2
GS1027	*Epilobium palustre* L.	Onagraceae	3476	1198646	523/128	56	30	3
GS1028	*Laserpitium latifolium* L.	Apiaceae	6167	1555461	258/125	75	7	3
GS1029	*Stellaria alsine Grimm*	Caryophyllaceae	2141	668566	374/125	47	24	3
GS1030	*Rumex obtusifolius* L.	Polygonaceae	588	307466	781/211	77	6	3
GS1031	*Galium rubrum* L.	Rubiaceae	2150	647899	289/160	73	6	4
GS1032	*Spiraea japonica* L.	Rosaceae	925	520774	780/227	78	6	1
GS1033	*Trichophorum cespitosum* (L.) Hartm.	Cyperaceae	818	398760	756/199	63	5	3
GS1034	*Carex pauciflora Lightf.*	Cyperaceae	224	190543	2638/276	63	4	2
GS1035	*Drosera rotundifolia* L.	Droseraceae	4097	1048550	258/128	14	5	3
GS1036	*Carex paupercula Michx.* subsp. *irrigua Löve*	Cyperaceae	407	261263	1443/217	70	4	3
GS1037	*Molinia caerulea* (L.) Moench	Poaceae	3982	1042723	272/125	74	3	2
GS1038	*Viola palustris* L.	Violaceae	7944	2262313	323/132	70	19	3
GS1039	*Luzula sudetica* (Willd.) Schult.	Juncaceae	4616	1029057	223/125	7	2	2
GS1041	*Athyrium distentifolium* Tausch ex Opiz	Athyriaceae	5187	1735601	396/151	58	1	3
GS1042	*Athyrium filix-femina* (L.) Roth	Athyriaceae	5874	2041099	421/163	44	1	3
GS1043	*Ligusticum mutellina* (L.) Crantz	Apiaceae	7160	1883759	278/131	76	9	3
GS1044	*Coeloglossum viride* (L.) Hartm.	Orchidaceae	14278	3295853	234/126	78	9	3
GS1045	*Doronicum austriacum* Jacq.	Asteraceae	5885	1506227	265/135	44	6	3
GS1046	*Lactuca alpina* (L.) A. Gray	Asteraceae	6742	1647306	245/140	52	4	3
GS1047	*Carex canescens* L.	Cyperaceae	900	271562	325/150	51	2	1
GS1048	*Ranunculus villarsii* DC.	Ranunculaceae	35707	6898784	193/125	78	6	3
GS1049	*Veronica alpina* L.	Plantaginaceae	3854	1334896	474/136	81	36	3
GS1050	*Cerastium fontanum Baumg.*	Caryophyllaceae	6466	1774318	287/134	61	29	2
GS1051	*Taraxacum palustre (aggregato)*	Asteraceae	10447	2760761	275/135	76	20	2
GS1052	*Cirsium alsophilum* (Pollini) Soldano	Asteraceae	5610	1665279	324/137	76	23	2
GS1053	*Bistorta officinalis Delarbre*	Polygonaceae	7365	1965040	269/151	54	8	4
GS1055	*Kalmia procumbens* (L.) Gift, Kron et Stevens ex Galasso, Banfi et F. Conti	Ericaceae	342	135839	435/203	31	1	3
GS1057	*Diphasiastrum alpinum*	Lycopodiaceae	6379	1558842	248/125	11	1	2

	(L.) Holub							
GS1058	*Empetrum hermaphroditum Hagerup*	Ericaceae	2548	744464	280/156	34	5	4
GS1059	*Lycopodium clavatum* L.	Lycopodiaceae	3993	1101624	288/135	22	2	2
GS1060	*Hieracium umbrosum* Jord.	Asteraceae	6817	1696282	254/128	78	6	3
GS1061	*Polystichum lonchitis* (L.) Roth	Dryopteridaceae	12505	6611722	695/246	75	0	3
GS1062	*Senecio germanicus* Wallr.	Asteraceae	3052	765457	245/151	57	3	4
GS1063	*Saussurea discolor* (Willd.) DC.	Asteraceae	14377	3434388	244/126	85	14	3
GS1064	*Hieracium amplexicaule* L.	Asteraceae	13431	3161388	243/126	78	7	4
GS1065	*Primula daonensis* (Leyb.) Leyb.	Primulaceae	6098	1747982	298/127	72	32	3
GS1067	*Dryopteris expansa* (C. Presl) Fraser-Jenk. et Jermy	Dryopteridaceae	8830	2542244	317/129	23	1	3
GS1068	*Agrostis schraderiana Bech.*	Poaceae	14668	3196199	223/125	79	5	3
GS1069	*Hieracium intybaceum* All.	Asteraceae	6555	1774211	283/141	85	6	4
GS1070	*Epilobium angustifolium* L.	Onagraceae	3373	978174	296/141	48	9	3
GS1072	*Cirsium erisithales* (Jacq.) Scop.	Asteraceae	6681	1857627	289/126	79	33	3
GS1073	*Alchemilla connivens Buser*	Rosaceae	443	240896	2159/168	72	6	3
GS1074	*Rhododendron hirsutum* L.	Ericaceae	2796	925462	322/198	43	7	3
GS1075	*Centaurea haynaldii Borbás subsp julica* (Hayek) Mayer	Asteraceae	23367	5632201	245/128	79	33	4
GS1076	*Adenostyles alliariae* (Gouan) A. Kern.	Asteraceae	7688	1846980	243/131	71	6	4
GS1078	*Epilobium alpestre* (Jacq.) Krock.	Onagraceae	3599	1201441	426/137	42	18	3
GS1079	*Saxifraga aizoides* L.	Saxifragaceae	1618	561825	399/158	70	6	3
GS1080	*Epilobium montanum* L.	Onagraceae	2655	880472	425/136	60	14	3
GS1081	*Stachys sylvatica* L.	Lamiaceae	6210	1536862	252/125	77	13	3
GS1082	*Calamagrostis varia* (Schrad.) Host	Poaceae	13589	3262115	235/128	71	11	1
GS1084	*Cytisus nigricans* L.	Fabaceae	5309	1455288	296/133	70	6	2
GS1085	*Trifolium medium* L. subsp. *medium*	Fabaceae	26414	7730720	296/195	57	5	3
GS1087	*Goodyera repens* (L.) R. Br. in W. T. Aiton	Orchidaceae	15944	4649041	289/194	76	7	2
GS1088	*Pimpinella saxifraga* L.	Apiaceae	3282	834742	239/189	57	5	3
GS1089	*Rosa inodora* Fr.	Rosaceae	1612	539559	356/150	80	7	3
GS1090	*Asperula purpurea* (L.) Ehrend.	Rubiaceae	223	167647	1279/283	60	6	1
GS1091	*Prunus avium* L. subsp. *avium*	Rosaceae	2196	691754	320/154	77	11	3
GS1092	*Asperula cynanchica* L.	Rubiaceae	1273	360404	285/164	42	4	3
GS1094	*Allium oleraceum* L.	Amaryllidaceae	20043	3937287	199/125	73	6	3
GS1095	*Dianthus seguieri* Vill.	Caryophyllaceae	11261	2632418	236/133	47	6	4
GS1096	*Vicia cracca* L.	Fabaceae	6302	1571647	254/130	29	5	3

GS1097	*Daucus carota* L.	Apiaceae	7864	1927382	249/125	78	12	2
GS1099	*Elytrigia repens* (L.) Desv.	Poaceae	9594	2070811	217/125	79	3	2
GS1100	*Clematis vitalba* L.	Ranunculaceae	3080	813467	264/126	76	5	3
GS1101	*Eupatorium cannabinum* L.	Asteraceae	8659	2103691	245/125	79	10	3
GS1102	*Rosa corymbifera Borkh.*	Rosaceae	1160	415648	417/157	70	5	3
GS1103	*Carduus acanthoides* L.	Asteraceae	3980	1094031	281/134	78	7	5
GS1104	*Centaurea bracteata* Scop.	Asteraceae	10014	2604688	273/137	79	8	4
GS1105	*Allium carinatum* L.	Amaryllidaceae	20220	4015073	198/125	79	6	3
GS1106	*Anthyllis vulneraria* L.	Fabaceae	13278	3378980	250/125	73	36	2
GS1107	*Rosa villosa* L.	Rosaceae	1512	508169	363/147	77	7	3
GS1108	*Epipactis helleborine* (L.) Crantz	Orchidaceae	10742	2427619	225/125	71	5	3
GS1109	*Anthericum ramosum* L.	Asparagaceae	17518	3910923	229/125	68	7	3
GS1110	*Knautia dipsacifolia Kreutzer*	Dipsacaceae	2667	752700	290/139	44	6	3
GS1111	*Ononis spinosa* L.	Fabaceae	10918	2810132	270/128	52	21	3
GS1112	*Calamagrostis epigejos* (L.) Roth	Poaceae	5057	1406892	304/130	76	2	4
GS1113	*Epilobium ciliatum* Raf.	Onagraceae	3172	964543	356/132	70	11	3
GS1114	*Sinapis arvensis* L. subsp. *arvensis*	Brassicaceae	1728	528011	336/125	80	6	3
GS1115	*Veronica anagallis-aquatica* L. subsp. *anagallis-aquatica*	Plantaginaceae	1637	640671	479/168	49	11	0
GS1116	*Juncus bufonius* L.	Juncaceae	1685	428320	239/191	14	1	2
GS1117	*Fallopia convolvulus* (L.) Á. Löve	Polygonaceae	6446	4904127	1230/328	66	6	3
GS1118	*Salix cinerea* L.	Salicaceae	2307	663313	276/132	74	7	3
GS1119	*Agrostis stolonifera* L.	Poaceae	6632	1630082	257/125	76	3	2
GS1120	*Gentiana pneumonanthe* L. subsp. *pneumonanthe*	Gentianaceae	3634	951186	270/131	77	6	3
GS1121	*Salix pentandra* L.	Salicaceae	605	225277	669/125	59	6	2
GS1122	*Salix rosmarinifolia* L.	Salicaceae	1546	1065699	1808/236	73	32	5
GS1123	*Thalictrum lucidum* L.	Ranunculaceae	9729	2927659	307/139	78	30	4
GS1124	*Mentha arvensis* L.	Lamiaceae	2298	801843	439/149	74	9	2
GS1125	*Epipactis palustris* (L.) Crantz	Orchidaceae	4381	1093039	248/126	75	5	3
GS1126	*Carex elata* All.	Cyperaceae	1059	432058	514/180	67	4	2
GS1127	*Succisa pratensis* Moench	Dipsacaceae	18444	4363198	241/125	59	17	0
GS1128	*Galium uliginosum* L.	Rubiaceae	13818	3555904	255/129	72	35	4
GS1129	*Galium palustre* L.	Rubiaceae	3894	1219704	349/140	52	33	4
GS1130	*Blysmus compressus* (L.) Panz. ex Link	Cyperaceae	381	230656	3392/188	60	5	2
GS1132	*Scabiosa columbaria* L.	Dipsacaceae	14020	3338730	235/125	54	37	1
GS1133	*Salix triandra* L.	Salicaceae	4772	1501635	374/138	66	16	4
GS1134	*Epilobium hirsutum* L.	Onagraceae	3657	1190733	440/129	46	19	3
GS1135	*Stachys officinalis* (L.) Trevis.	Lamiaceae	2674	576778	226/127	4	0	1
GS1137	*Centaurea jacea* L.	Asteraceae	8383	2265573	284/142	80	7	2
GS1138	*Aster amellus* L.	Asteraceae	13739	3349415	250/134	77	9	3
GS1139	*Parnassia palustris* L. subsp. *palustris*	Celastraceae	7561	2201589	272/155	45	37	2

GS1140	*Carex capillaris* L.	Cyperaceae	748	244016	369/164	47	1	3
GS1141	*Juncus alpino-articulatus* Chaix	Juncaceae	2697	768133	305/135	40	5	1
GS1142	*Epilobium parviflorum* Schreb.	Onagraceae	4558	1430090	390/127	67	26	3
GS1143	*Sorbus aria* (L.) Crantz	Rosaceae	2344	650416	279/145	58	8	1
GS1144	*Nigritella nigra* (L.) Rchb. fil.	Orchidaceae	10584	2536035	245/130	76	4	3
GS1145	*Senecio doronicum* (L.) L.	Asteraceae	3750	1045392	283/140	79	6	3
GS1146	*Luzula spicata* (L.) DC.	Juncaceae	5544	1368479	255/125	15	2	4
GS1147	*Dryopteris villarii* (Bellardi) Woyn. et Thell. subsp. *villarii*	Dryopteridaceae	16605	4470558	288/125	56	1	3
GS1148	*Thesium pyrenaicum* Pourr.	Thesiaceae	69143	17582394	245/191	49	36	1
GS1149	*Leucanthemum heterophyllum* (Willd.) DC.	Asteraceae	6564	1563004	242/125	82	5	3
GS1150	*Helictotrichon parlatorei* (Woods) Pilg.	Poaceae	15426	3711730	251/126	78	3	4
GS1151	*Trisetaria argentea* (Vill.) Banfi et Soldano	Poaceae	12152	2732799	227/125	75	5	3
GS1152	*Sedum atratum* L.	Crassulaceae	3404	1026408	301/154	69	16	2
GS1153	*Campanula cochleariifolia* Lam.	Campanulaceae	19363	4268286	222/125	66	12	3
GS1154	*Orobanche reticulata* Wallr.	Orobanchaceae	10562	3676075	492/137	24	32	2
GS1155	*Heracleum sphondylium* L.	Apiaceae	10272	2297197	221/125	80	8	3
GS1156	*Helianthemum nummularium* (L.) Mill.	Cistaceae	8687	1846414	213/125	20	6	4
GS1157	*Gymnadenia odoratissima* (L.) Rich.	Orchidaceae	7109	1863057	269/137	76	4	3
GS1159	*Ranunculus thora* L.	Ranunculaceae	19013	4823659	261/125	76	28	4
GS1160	*Asperula aristata* L. fil.	Rubiaceae	1014	334046	368/161	51	2	4
GS1161	*Hieracium bupleuroides* C. C. Gmel.	Asteraceae	18900	4525646	246/126	78	22	4
GS1163	*Leontodon hispidus* L.	Asteraceae	10422	2612432	246/125	77	35	4
GS1164	*Stachys recta* L.	Lamiaceae	8013	1952945	246/125	82	19	2
GS1165	*Rhaponticum scariosum* Lam.	Asteraceae	7313	2073130	314/126	80	19	4
GS1166	*Orobanche alba* Stephan ex Willd.	Orobanchaceae	10497	3669758	487/148	9	16	2
GS1167	*Scrophularia hoppii Koch*	Scrophulariaceae	391	235265	3879/206	73	5	3
GS1168	*Erigeron glabratus* Bluff et Fingerh.	Asteraceae	9797	2341965	245/126	73	12	4
GS1169	*Phyteuma zahlbruckneri Vest*	Campanulaceae	3408	939806	296/127	71	3	3
GS1170	*Doronicum grandiflorum* Lam.	Asteraceae	15325	3503674	233/125	78	6	2
GS1171	*Salix reticulata* L.	Salicaceae	1914	575521	292/140	71	8	3
GS1172	*Alchemilla colorata Buser*	Rosaceae	647	274986	648/164	80	6	4
GS1175	*Campanula scheuchzeri* Vill.	Campanulaceae	7227	1725479	239/128	59	3	3
GS1177	*Cystopteris alpina* (Lam.) Desv.	Cystopteridaceae	5292	1726232	396/143	16	1	3
GS1178	*Cerastium latifolium* L.	Caryophyllaceae	3554	1108912	340/153	48	12	3

GS1180	*Dryas octopetala* L. subsp. *octopetala*	Rosaceae	1276	664917	780/217	75	21	0
GS1181	*Salix retusa* L.	Salicaceae	1469	489234	328/170	73	8	4
GS1182	*Alchemilla fallax Buser*	Rosaceae	2529	705248	277/130	79	7	5
GS1183	*Silene acaulis* (L.) Jacq.	Caryophyllaceae	14836	3777499	264/132	70	8	3
GS1184	*Rhodothamnus chamaecistus* (L.) Rchb.	Ericaceae	318	218583	2440/221	39	5	4
GS1185	*Festuca quadriflora Honck.*	Poaceae	7533	1970582	275/134	72	4	3
GS1186	*Salix alpina* Scop.	Salicaceae	1660	515742	303/137	74	6	4
GS1187	*Adenostyles glabra* (Mill.) DC.	Asteraceae	13977	3135983	225/125	78	9	4
GS1188	*Saxifraga caesia* L.	Saxifragaceae	2247	722626	364/158	60	7	4
GS1189	*Leontopodium alpinum Cass.*	Asteraceae	11894	3204019	283/137	78	12	3
GS1190	*Bupleurum petraeum* L.	Apiaceae	1884	585315	340/132	73	6	4
GS1191	*Aquilegia brauneana (Hoppe) Ják.*	Ranunculaceae	107359	26906693	244/188	65	32	3
GS1192	*Alchemilla strigosula Buser*	Rosaceae	1311	449119	390/140	79	6	3
GS1193	*Elymus caninus* (L.) L.	Poaceae	10696	2469129	234/125	76	17	3
GS1194	*Galeopsis pubescens Besser* subsp. *pubescens*	Lamiaceae	8359	2170172	275/125	76	20	3
GS1195	*Carex parviflora Host*	Cyperaceae	652	285119	594/197	59	2	3
GS1196	*Veronica aphylla L.*	Plantaginaceae	5383	1715293	375/128	69	36	2
GS1198	*Alchemilla fissa* Günther et Schummel	Rosaceae	835	326836	527/150	77	6	3
GS1199	*Hieracium villosum* Jacq.	Asteraceae	13506	3300880	252/132	77	7	4

By contrast, the average number of CDS for chloroplast and mitochondria and of ribosomal RNA genes were 64.20, 14.14 and 2.84, respectively (Figure 3.8). Due to the number of CDS in chloroplast (about 80) and mitochondia (around 40-50) in previous studies (Raju et al., 2016; Kubo et al., 2008), it is clear that the sampling of the gene space was very successful for chloroplast CDS (80.3%), while relatively incomplete for mitochondrial genes (31.4%). As the CDS were extracted based on homology to the most closely related genomes in Genbank, it is possible that this difference may stem at least in part from the relatively few mitochondrial genomes available as reference. As the species considered in this study have a broad taxonomic span, it is possible that the much higher taxonomic representation of plastomes makes their annotation more efficient, while the sparser taxonomic sampling for mitogenomes could have caused the use of suboptimal (too distantly related) reference genomes. The much higher sampling of plastome CDS, however, is in line with the type of tissues used for DNA extraction, which were selected to maximize the amount of plastidial DNA over mitochondrial and nuclear genomes (Petit et al., 2005). rRNA genes were sampled even more efficiently than plastid genes, but as they were not the primary focus of this work, the assembly strategy here employed was not designed to preserve the allelic variability of these genes known from previous reports (Arrigoni et al., 2017) .

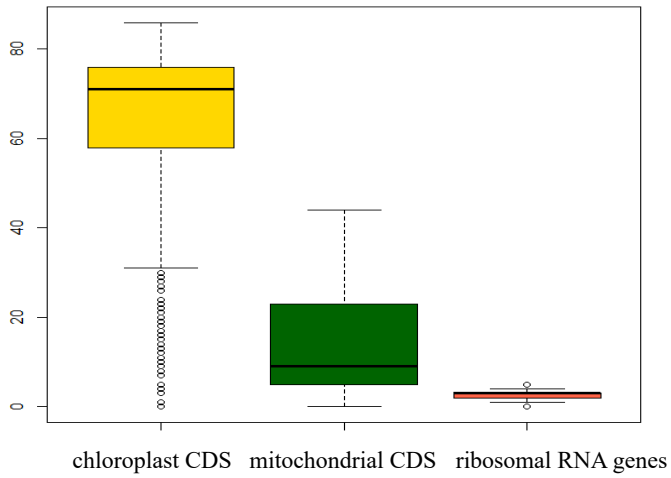

Figure 3.8: Gene number of chloroplast, mitochondrial and ribosomal RNA in 1037 angiosperms species. The average number of chloroplast CDS is 64.20, that of mitochondrial CDS is 14.14 and that of ribosomal RNA genes is 2.84.

3.3.2 Selected families in angiosperms

Despite the good sampling of the plastome gene space, the left tail distribution of the gene number among species could negatively affect the phylogenetic and molecular evolutionary analyses by inflating the number of missing data. Additionally, the inherent dishomogeneity of the taxonomic representation of the studied flora and the incomplete sampling suggested me to focus all subsequent analyses on the most represented families. The top 11 families, chosen because encompassing more than 25 species, included 8 eudicots (Apiaceae, Asteraceae, Caryophyllaceae, Fabaceae, Lamiaceae, Plantaginaceae, Ranunculaceae and Rosaceae) and 3 monocots (Cyperaceae, Orchidaceae and Poaceae) clades (Table 3.8). From this table, the lowest average number of CDS was in the Cyperaceae family (69) while the largest average number of CDS was in the Asteraceae family (79), and the total number of alignments was 838 (family/gene combinations). The distribution of genes across families is shown in Figure 3.9. This approach on the one hand simplified the dataset (constituted now by 578 taxa and 38,778 genes) and on the other hand made it more uniform in terms of gene space completeness and evenness. As expected by the high representation of the gene space and by the functional composition of the chloroplast gene space, we found that photosystem II had the largest number of genes in all the families, followed by NADH dehydrogenase and ribosomal subunit genes. Furthermore, there were 7 classes of genes that had only one representative in each family, such as membrane protein, maturase etc.

Table 3.8: Summary of 11 families for evolutionary analyses

Families	Number of species	CDS alignments	BUSTED	FUBAR	MEME	aBSREL	RELAX
Apiaceae	32	77	+	+	+	+	-
Caryophyllaceae	47	76	+	+	+	+	-
Lamiaceae	32	78	+	+	+	+	-
Plantaginaceae	31	77	+	+	+	+	-
Ranunculaceae	36	77	+	+	+	+	-
Asteraceae	117	79	+	+	+	+	+
Fabaceae	67	75	+	+	+	+	+
Rosaceae	57	77	+	+	+	+	+
Cyperaceae	49	69	+	+	+	+	+
Poaceae	82	76	+	+	+	+	+
Orchidaceae	29	77	+	+	+	+	-

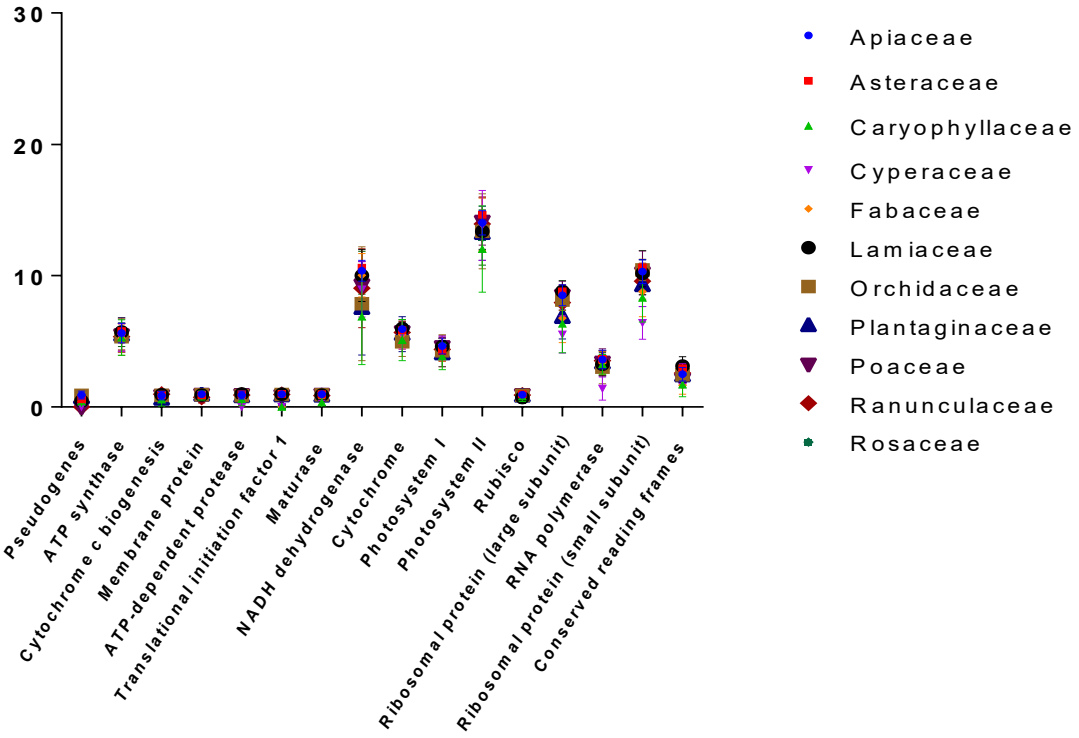

Figure 3.9: The classification of chloroplast CDS in 11 families by function. Different colors correspond to different families. The Y-axis reports the number of genes in each of the functional classes listed on the X-axis.

3.3.3 Natural selection analyses

3.3.3.1 Gene-wide selection analyses

After validation and refinement of the working set of taxa/genes, the in-depth analyses of the patterns of molecular selection in chloroplasts were carried out through different packages of the HyPhy program. As the chloroplast genes are inherited maternally, the likelihood of recombination between paralogous copies with different evolutionary histories is limited to the few taxa where inheritance is biparental, e.g. some *Pergonium* species and conifers (Petit et al., 2005). Therefore, recombination tests like those implemented in single breakpoint (SBP) or genetic algorithm recombination detection (GARD) programs which are also extremely time consuming (Pond et al., 2006) were skipped in this work. As a first step, it was tested whether the evolutionary patterns of whole-gene selection are conserved among the selected families or not. BUSTED is a program of the HyPhy package that performs a gene-wide identification of selection based on a branch-site unrestricted statistical test for episodic diversification (Murrell et al., 2015). As all tests with relatively few parameters and assumptions, BUSTED is expected to provide a relatively robust estimation of the number of genes under selection in each family. The results of the BUSTED analyses are shown in Figure 3.10, where the blue bars represent the number of families in which each gene showed signatures of positive selection. A total of 25 genes across the 11 families were under positive selection according to BUSTED. The *rbcL* gene was the only gene identified as under positive selection in the large majority of the families considered (8 out of 11) after correcting for multiple tests by p-value (FDR=5%). Other genes repeatedly identified as under positive selection were found at most in 2 or 3 families, like the *ccsA, rpoB, ycf1* genes in 3 families, and the *atpA, matK, ndhF, psaA, psbB, rpoC2,* and *ycf2* genes in 2 families. Besides, there were 14 genes detected under positive selection only in one family.

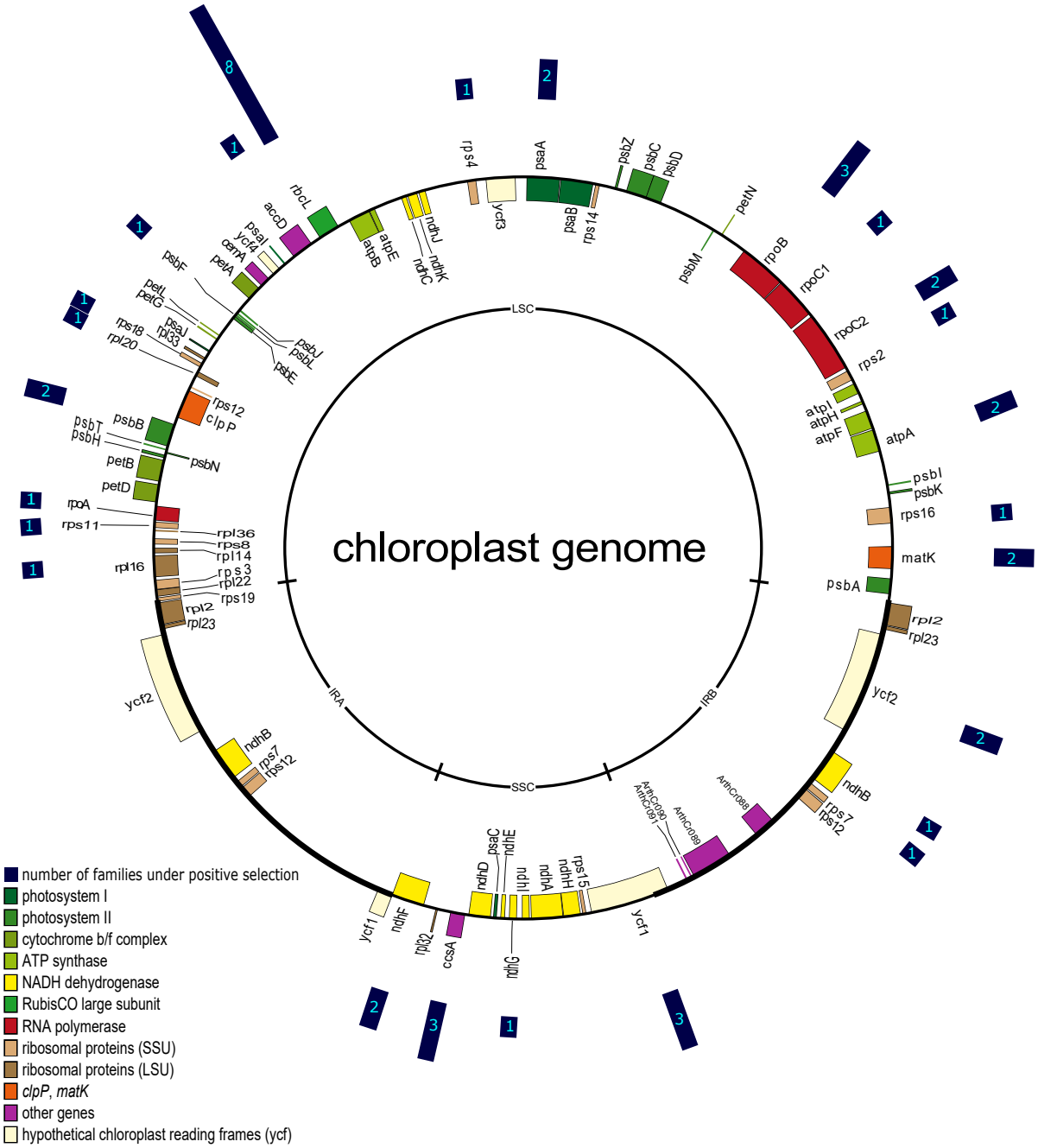

Figure 3.10: Numbers of families for chloroplast CDS under positive selection across families. This figure uses *Arabidopsis thaliana* as reference genome to display the results of gene wide selection. The inner-circle indicates the four major regions of the chloroplast genome, mid-circle illustrates the function of genes by different colors, and the outside-circle presentes number of families in each gene under positive selection as detected by BUSTED program.

In order to assess to what degree the above assumptions can be transgressed in the current dataset, basic statistics were calculated on the number of taxa, alignment lengths and average number of differences among all alignments. Summarizing the mean number of differences per gene and alignment length across families, the distribution of the index is approximately normal and the value for *rbcL* (2.7) falls in the bulk of the distribution, very close to the average value (3.1). To check in more detail whether the polymorphism level of the different alignments could affect the rate of whole gene positive selection detection and explain the high number of hits for *rbcL*, whether the number of hits expressed as a function of the absolute number of differences per alignment were related to each other was further tested. The plot was interpolated with different functions (linear, exponential, logarithmic, polynomial of order 2 to 6) and the best fit (R^2=0.3004) was obtained with a 3rd order polynomial function, whose equation is $y= 2^{-07}x^3 – 0.0002x^2 + 0.0356x – 0.1104$ (Figure 3.11). *rbcL* is clearly separated from the other genes along the y-dimension, while *ycf1* is the gene with the highest number of differences per alignment. Whether they could be identified as outliers was checked by the interquartile range (IQR) test. When applied to the vector of alignment differences, six genes were found to be distribution outliers (three major outliers: *ycf1*, *ndhF* and *rpoC2*; and three minor outliers: *rpoB*, *matK* and *ycf2*). The *rbcL* gene was not an outlier in this dimension, but it was the only major outlier in the dimension of the number of BUSTED hits per family, indicating that the high number detected by BUSTED where *rbcL* was identified among the positively selected genes in different families does not depend from either alignment length or number of differences, but may be the direct consequence of recurrent selection acting specifically on this gene.

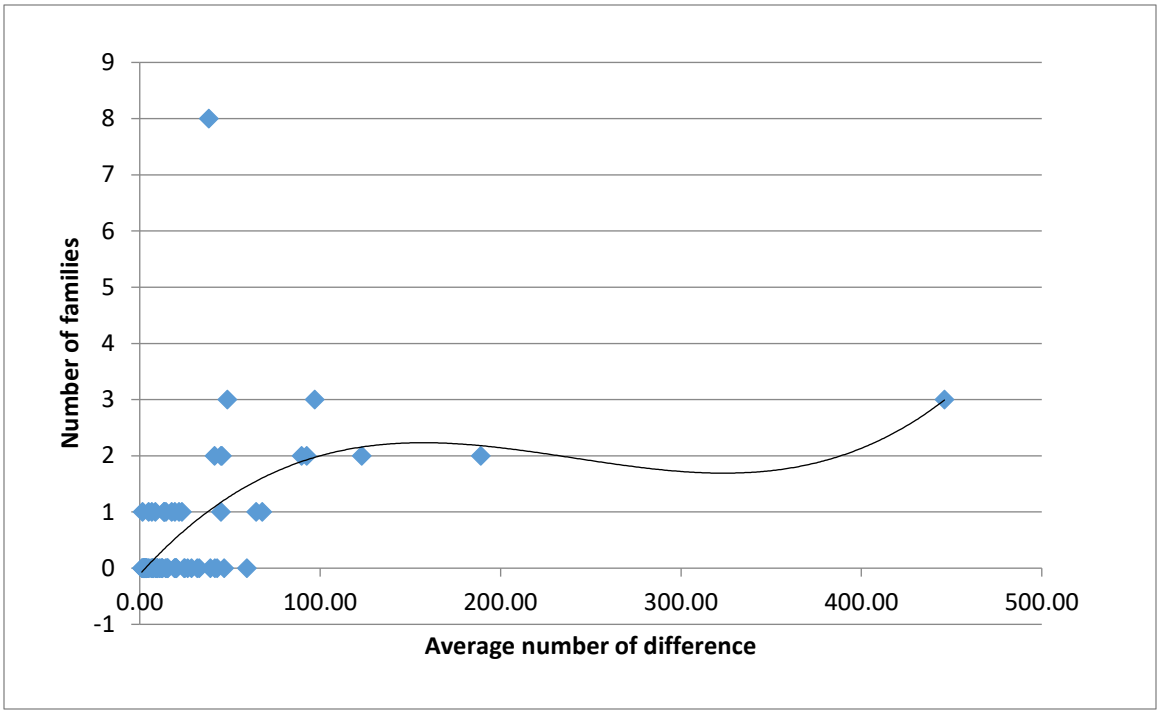

Figure 3.11: The correlation between the rate of gene-wide positive selection and the polymorphism level of the different alignments. The Y-axis reports the number of families in each of positively selected gene with average number of differences per gene and alignment length across families listed on the X-axis.

3.3.3.2 Site level selection analyses

In addition to the analyses aimed at detecting gene-wide positive selection, site-level and branch-level selective patterns were also analyzed in the same families. The MEME program in HyPhy is a particularly popular and useful method designed to detect selection at the level of individual sites by different codon substitution models (Murrell et al., 2012a). After FDR correction of MEME p-value results at 5%, only 4 sites were identified as under positive selection in 11 families. They were codon site number 956 in the *rpoC2* gene of the Caryophyllaceae family, codon site 197 in the *ccsA* gene of the Fabaceae family, codon site 174 in the *psaA* gene and codon site 247 in the *rbcL* gene of the Rosaceae family. Actually, site 956 of *rpoC2* was represented by phenylalanine (TTT; 19 sequences) and tyrosine (TAT; 13 sequences), two aromatic amino acids with relatively similar physico-chemical properties, with some substitutions with asparagine (AAA; 5 sequences) or valine (GTT; 1 sequence). The site 197 of *ccsA* was mainly serine (TCT), with only 3 substitutions with tyrosine (TAT; 1 sequence) or isoleucine (ATA; 2 sequences). The site 174 of *psaA* was serine (TCT; 30 sequences) and glycine (GGT; 15

sequences), with just an additional sequence with an alanine mutation (GCT). The site 247 of *rbcL* was cysteine (TGC; 41 sequences), valine (GTC; 7 sequences) and asparagine (AAC; 3 sequences). Compared with BUSTED results, all of these 4 genes were detected as under positive selection at both gene-wide and site levels, indicating that the stringent multiple test correction was effective in reducing the number of false positives and focusing only on the most clear-cut cases of sites under positive selection. Structural modelling of RbcL and PsaA proteins on the available crystal structures from the PDB database was carried out, in the hope that it could provide further insights into the possible role of the positions under positive selection.

3.3.3.3 Branch level selection analyses

The branch level method is based on adaptive branch-site random effects likelihood (aBSREL), which is a powerful and popular approach for detecting signatures of natural selection from molecular data. The aBSREL method infers the optimal number of rate categories to be used for each branch through the small-sample AIC. Table 3.9 summarizes the 46 branches which were detected as under positive selection by multiple test correction (FDR=5%) among 838 alignments in 11 families. A total of 60 alignments were positive either in BUSTED (45) or aBSREL (36). Among them, 21 (about one third) of the alignments were consistently identified as positive by both programs. When considering only the genes identified as positively selected by both programs, the majority of the positive genes detected by each program were detected also by other programs (17 out of 25 or 20 for BUSTED or aBSREL, respectively).

Table 3.9: Branches under positive selection in alignments of 11 families by aBSREL

Family	Branch	P-value	FDR
Apiaceae	psbB_GS0300	2.99012E-05	0.048779
Asteraceae	ndhB_Node189	6.27554E-06	0.014717
Asteraceae	ndhF_GS1062	4.47755E-07	0.001867
Asteraceae	rbcL_GS0968	7.42969E-06	0.016895
Caryophyllaceae	ndhF_GS1001	5.38036E-06	0.013458
Caryophyllaceae	rpoC1_GS0774	4.58039E-09	3.82E-05
Caryophyllaceae	rpoC2_GS0798	3.16414E-15	1.19E-10
Caryophyllaceae	rpoC2_Node3	1.51713E-09	1.63E-05
Cyperaceae	atpA_GS1033	6.25867E-06	0.014717
Cyperaceae	rpoB_GS0005	8.87248E-06	0.017521
Cyperaceae	rpoB_GS0129	2.38548E-05	0.041631
Cyperaceae	rpoB_GS0133	8.85049E-06	0.017521
Cyperaceae	rpoC1_GS1140	1.46686E-06	0.004234
Cyperaceae	rps2_Node4	3.10499E-07	0.001446

Cyperaceae	rps4_Node4	2.41184E-07	0.001221
Fabaceae	ccsA_Node14	1.25881E-06	0.003779
Fabaceae	ccsA_Node37	1.60811E-05	0.030169
Fabaceae	ccsA_Node40	2.02056E-05	0.036102
Fabaceae	ndhF_Node79	2.64718E-06	0.00685
Fabaceae	rpoB_GS0058	1.14177E-06	0.003633
Fabaceae	rpoB_GS0263	9.92116E-07	0.003384
Fabaceae	rpoC1_GS0263	3.36211E-09	3.15E-05
Fabaceae	ycf1_Node12	2.23264E-07	0.001221
Fabaceae	ycf1_Node4	7.88588E-06	0.017045
Fabaceae	ycf1_Node9	1.1419E-07	0.000779
Lamiaceae	ccsA_Node48	4.23054E-11	6.35E-07
Lamiaceae	ndhG_GS0341	1.63078E-07	0.00102
Orchidaceae	atpF_Node4	8.55913E-06	0.017521
Orchidaceae	rpl16_GS0918	5.13691E-07	0.001927
Orchidaceae	rpoA_Node6	2.44094E-07	0.001221
Orchidaceae	rps16_Node26	4.98399E-07	0.001927
Orchidaceae	ycf2_Node2	2.2603E-11	4.24E-07
Orchidaceae	ycf2_Node5	7.37577E-13	1.84E-08
Plantaginaceae	clpP_GS0410	7.94993E-06	0.017045
Plantaginaceae	ndhF_GS0990	1.01129E-07	0.000759
Plantaginaceae	ndhK_GS0420	1.20982E-09	1.51E-05
Plantaginaceae	rbcL_Node2	2.95707E-05	0.048779
Poaceae	ndhF_GS0469	2.55688E-06	0.00685
Poaceae	ndhF_Node70	6.65191E-07	0.002377
Poaceae	psbB_Node59	3.2754E-07	0.001446
Poaceae	rpl16_Node56	2.74424E-05	0.046803
Ranunclaceae	rpoB_Node52	1.78063E-05	0.032591
Ranunclaceae	ycf2_GS0326	1.16193E-06	0.003633
Rosaceae	rpoB_GS0827	2.33887E-06	0.006501
Rosaceae	ycf1_GS0861	9.16984E-06	0.017644
Rosaceae	ycf2_Node52	7.21645E-16	5.42E-11

3.3.3.4 Relaxed selection analyses

It is known from literature that altitude constitute a strong barrier to the colonization of new taxa, a process that is called ecological filtration. In the Trentino region, former studies have found clear evidences of ecological filtration mediated by altitude adaptation (Marini et al., 2008), suggesting that it could be possible to associate the patterns of positive selection with this important feature of the landscape. To test this hypothesis, 5 families (Asteraceae, Cyperaceae, Fabaceae, Poaceae and Rosaceae) among the extended

11 families dataset, which had a sufficient number of species distributed at different altitude, were chosen for relaxed selection analysis. Altitude distribution is usually normally distributed, with the average of the distribution representing the optimal altitude preference for the species. Given the sometimes high standard deviation of the distribution (or, in other terms, the broad altitude distribution range of some species), the species were divided in each of the 5 families in roughly 3 groups, representing low, medium and high alititude species based on their average altitude preference. Excluding middle altitude species from analysis, the sensitivity of the detection of differences was maximized between average omega values at the extremities of the distribution (low and high groups), to avoid the loss of power associated to species close to the middle of the altitude range. The summary of the number of test (high altitude), reference (low altitude) and unassigned (middle altitude) taxa for each family is shown in Figure 3.12. Thus, RELAX tested whether selection was relaxed or intensified on the subset of high-altitude test branches compared with the subset of low-altitude reference branches.

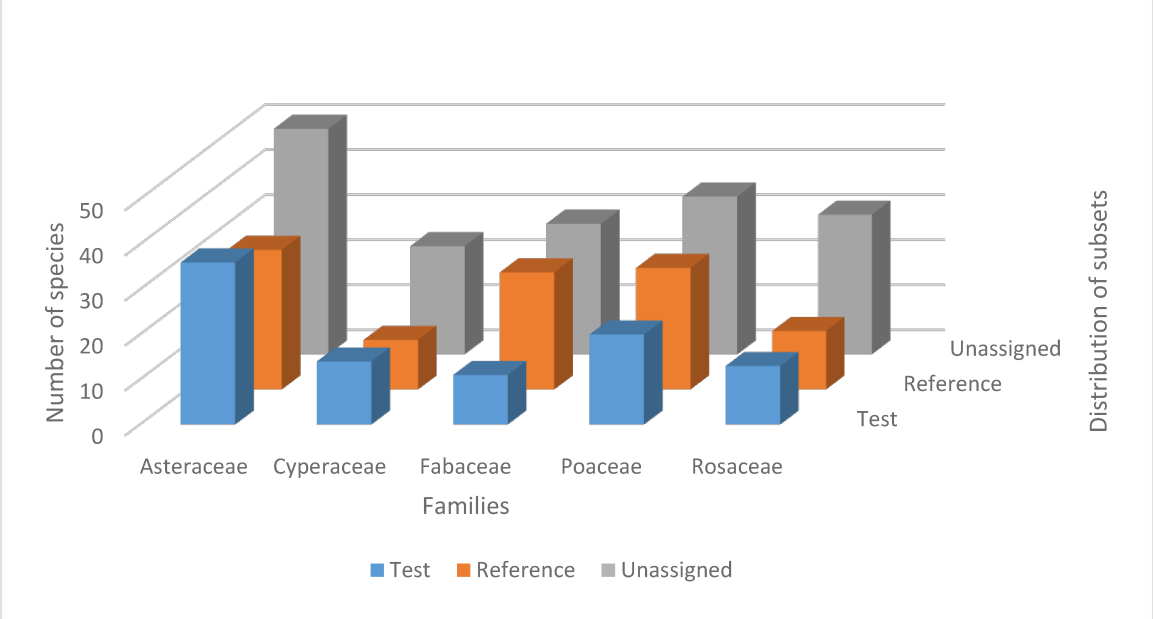

Figure 3.12: Distribution of subsets in 5 families by RELAX analyses. Blue color illustrates a subset of test species (high altitude) in each family, orange color presentes a subset of reference species (low altitude) in each family and gray color demonstrates the species which are unassigned (middle altitude) in RELAX.

Based on RELAX analyses in 376 alignments out of 5 families, 4 genes (3 genes when using species trees) were identified as under positive selection with FDR = 5% (Table 3.10), two in the Asteraceae family (*accD* and *rbcL*), and two in the Fabaceae family (*rpoC1* and *rps18*). The relaxation coefficients of the top three genes were less than 1, indicating relaxed selection, while only the *rps18* gene showed increased selection

intensity (K >1). Besides, the analysis did not detect any effect of altitude on selection of genes from the other three families, indicating that the association between altitude and omega is lineage-specific.

Table 3.10: Relaxed-selection in altitude biological trait for 5 families

Gene	Family	Relaxation Coefficient (K)	Likelihood Ratio (LR)	P-value	FDR
accD	Asteraceae	4.11E-07	14.139387	0.00017	0.021219
rbcL	Asteraceae	0.659682099	28.784119	8.09E-08	3.03E-05
rpoC1	Fabaceae	0.482614874	17.046208	3.65E-05	6.84E-03
rps18	Fabaceae	8.64621987	12.039462	0.000521	0.048831

3.3.4 RNA editing analysis

In angiosperms (flowering plants), RNA editing was first recognized in the *rpl2* transcript of maize in 1991 as DNA and RNA sequences from the same genes were found to differ (Hoch et al., 1991). The presence of uridine (U) nucleotides in the RNA in positions corresponding to cytidine (C) nucleotides in the DNA were imputed to post-transcriptional C-to-U replacements in the RNA. Such RNA-editing processes are used as control checkpoints, can restore the function of the encoded protein, and create different proteins (Takenaka et al., 2013). As RNA editing can mislead the prediction of the mature peptide encoded by plastidial mRNAs, further tests were used to assess whether it could affect also the patterns of molecular evolution detected in this study. Thus, RNA editing sites in gene sequence alignments were predicted in PREPACT2 based on 10 plastome references (Table 3.11). The codon positions predicted to be affected by RNA editing were deleted from the alignments and the analyses of the different routines of the HyPhy package run again on the revised alignments. This search did not take into account editing sites outside of the protein-coding regions (5'-UTRs, 3'-UTRs and introns), the synonymous positions of codons, or tRNA/rRNA genes.

Table 3.11: Numbers of RNA-editing sites in gene alignments of 11 families

Gene	Apiaceae	Asteraceae	Caryophyllaceae	Cyperaceae	Fabaceae	Lamiaceae	Orchidaceae	Plantaginaceae	Poaceae	Ranunculaceae	Rosaceae
accD	6	11	6	0	3	2	6	5	0	5	7
atpA	3	2	3	1	4	2	4	3	3	7	6
atpB	1	0	1	2	2	0	1	2	3	3	1
atpF	0	0	0	0	3	1	3	3	0	2	1
atpI	0	2	2	1	3	1	3	1	0	3	1
clpP	1	2	1	0	6	1	3	4	1	3	1
matK	6	6	5	9	13	10	9	13	13	13	11
ndhA	7	9	4	4	5	3	5	2	6	8	4
ndhB	13	11	11	13	18	10	14	11	13	13	13

ndhC	0	0	1	3	2	0	4	1	2	2	3
ndhD	7	8	9	7	10	5	11	4	3	11	9
ndhE	0	0	2	0	1	1	0	1	0	0	1
ndhF	6	7	16	20	10	9	14	12	11	16	12
ndhG	2	0	2	1	3	0	3	3	1	4	2
ndhH	1	2	1	0	2	4	2	2	0	2	2
ndhK	2	1	5	3	1	1	3	7	5	2	2
petB	2	2	1	0	2	2	2	2	1	1	2
petD	0	1	1	0	1	0	1	1	0	0	0
petG	1	1	1	1	1	1	0	1	0	0	1
petL	1	1	1	0	0	0	2	0	0	2	1
psaI	2	0	1	0	0	1	1	1	0	1	1
psbB	0	1	0	2	0	1	1	0	2	0	0
psbD	0	1	0	1	0	0	0	1	1	1	0
psbE	1	1	0	0	1	1	0	1	0	0	1
psbF	1	1	1	1	1	0	1	1	0	1	1
psbJ	0	0	1	0	0	0	0	1	0	0	0
psbL	0	0	0	0	0	0	0	0	0	0	0
psbN	0	0	0	0	0	0	0	0	0	0	0
psbZ	0	1	0	0	0	0	0	0	0	0	0
rpl20	1	1	3	3	4	2	5	4	2	3	2
rpl23	0	0	0	0	2	1	0	1	5	1	2
rpl2	0	2	2	2	4	1	0	2	0	1	0
rpl32	0	0	0	0	2	2	0	1	0	0	0
rpoA	7	3	4	0	5	2	4	5	5	3	4
rpoB	11	4	11	15	23	9	10	12	13	9	10
rpoC1	4	6	9	8	12	4	7	2	8	8	2
rpoC2	12	5	9	8	24	13	9	18	21	15	13
rps14	2	2	1	2	3	3	2	3	2	2	2
rps16	0	0	0	2	1	1	1	0	0	3	2
rps18	1	2	3	0	7	2	1	2	2	2	2
rps2	1	1	3	7	7	1	3	3	2	3	2
rps3	0	2	0	0	5	1	2	4	2	4	0
rps8	0	0	1	1	3	1	2	1	2	3	0
ycf2	20	32	17	0	0	14	31	0	0	37	17
ycf3	0	0	0	3	0	1	3	1	2	2	0

The original gene alignment datasets were compared with the revised gene alignment datasets, from which the putative RNA-editing sites were deleted. Gene-wide detection of positive selction was then applied to assess whether RNA editing influenced the inferred molecular evolutionary patterns or not in this study. Figure 3.13 shows that predictions of genes under positive selection were influenced to a moderate extent by RNA editing, as

19 out of 25 genes detected by BUSTED program were identified as positive selected genes in both datasets. Especially, the *rbcL* gene was not affected at all, as it was basically RNA editing free. A previous study illustrated that the *rbcL* gene was detected as positively selected in almost 43 orders and 203 families in flowering plants (Kapralov and Filatov, 2007). Patterns of positive selection in 8 out of 11 families tested were observed in this study, supporting recurrent selection on *rbcL*. Widespread positive selection in the *rbcL* gene, therefore, has to be taken into account when this gene is used for phylogenetic reconstructions and evolutionary studies. Besides *rbcL*, little differences were observed for the other 6 genes predicted as under positive selection by BUSTED after FDR correction (*atpA, ndhF, psbB, psbF, rpoC2,* and *rpoA*). Noteworthy, these differences could be detected in just one family.

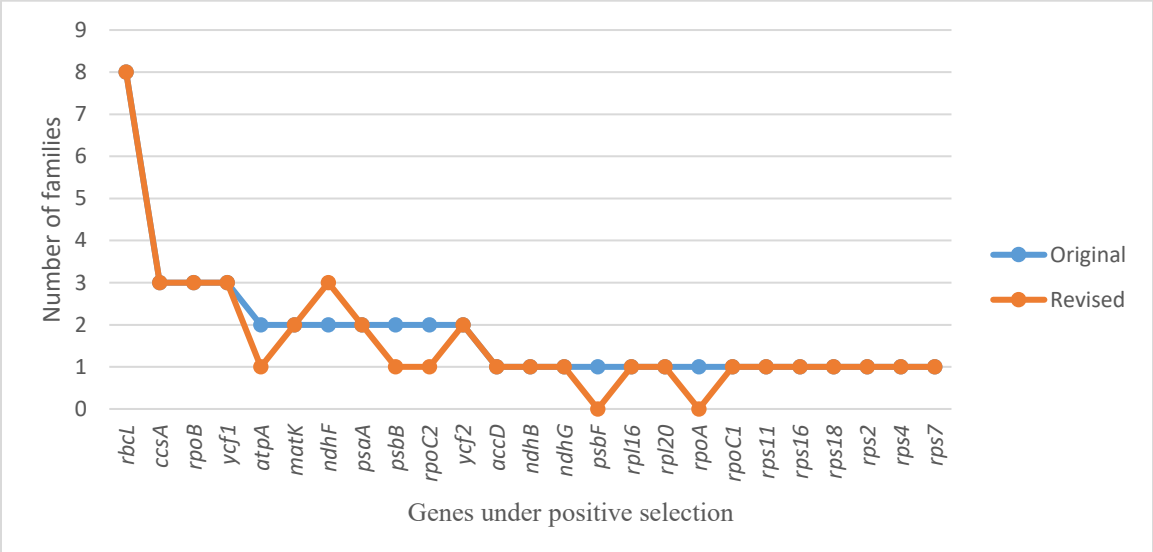

Figure 3.13: The different number of positively selected genes across families in original alignments and revised alignments. Blue color representes the genes under positive selection in original alignments while orange color representes the genes in revised alignments which RNA editing sites are deleted. This figure demonstrates that different numbers of families in chloroplast genes detected under positive selection in BUSTED.

Further test between original and revised alignment datasets in site level selection was performed by MEME and FUBAR programs in 11 families, as more than 90% of positive selection sites were the same in both of these datasets. On one hand, 1003 and 964 positive selection sites (p-value < 0.05) were obtained in original and revised datasets by the MEME program, respectively. And among them, 939 sites were supported in both of these datasets, whereas 64 sites and 25 sites were unique in original and revised datasets, respectively. Additionally, 27 sites were predicted as RNA editing sites in original dataset. On the other hand, 285 and 264 sites (probability > 0.95) were detected as under positive selection by the FUBAR program for original and revised datasets, respectively. The

supported sites in both of cases were 262, while 12 sites were predicted as RNA editing sites in the original dataset.

Due to the sensitivity of methods in natural selection analysis, a larger dataset of positive selection sites were obtained in MEME than in FUBAR. It is commonly known that MEME is capable of identifying instances of both episodic and pervasive positive selection at the level of an individual site. Murrell and colleagues demonstrated the superior performance of MEME over older models under a broad range of scenarios through empirical and simulated data (Murrell et al., 2012b). Episodic selection was concluded to be widespread, indicating that the number of sites experiencing positive selection may have been largely underestimated. On the other hand, the main advantage of FUBAR is the extreme speed increase, as it does not take heterotachy into account. Therefore, positive selection sites were detected by combining MEME and FUBAR programs, as this could increase the sensitivity of the detection.

Finally, branch level selection between original and revised datasets was tested by the aBSREL program. About 8 out of 11 families had the same branches identified under positive selection, and only Apiaceae, Cyperaceae and Fabaceae had several different branches detected as under positive selection. In Apiaceae, only the psbB_GS0300 branch was predicted to be under positive selection in the original dataset. This situation also occurred in Fabaceae, as only rpoB_GS0263 branch was under positive selection in original dataset. Additionally three branch differences were observed in Cyperaceae, for instance, rpoB_GS0129 and rpoB_GS0133 branches were predicted in revised dataset while rps2_Node1 branch was in original dataset. Besides these differential branches, the other branches were the same. Therefore, the above analyses of RNA editing at gene wide, site and branch level across 11 families indicated that RNA editing had little influence on selection analyses in this study.

4 Discussion

4.1 Phylogenetic reconstruction of the European *Aquilegia* rapid radiation by next-generation sequencing

4.1.1 Correlation between GC content and amino acid polymorphisms

It is generally known that chloroplast genomes of land plants are highly conserved. This conservation is considered to largely result from the evolutionary constraints associated with photosynthesis (Bungard, 2004). A previous study has shown that photosynthetic chloroplast genes in rice, tobacco and liverwort have higher GC content and higher GC preferences in the third position than other genes (Shimda and Sugiuro, 1991), and this conclusion is confirmed also by the analyses carried out in the *Aquilegia* genus. According to analysis of the GC content, different codon positions have obviously different proportion of GC content in *Aquilegia*, as the first position had the highest proportion of GC content (48%) while the third codon position had the lowest (32%). Knight et al. demonstrated that the patterns of codon usage are influenced by its composition bias (Knight et al., 2001), and codon positions with low GC content are more variable than others. For instance, the low GC content gene *matK* evolved approximately three times faster than *rbcL* in *Saxifragaceae* (Johnson and Soltis, 1994). Although the evolutionary rates of these two genes were not detected in the *Aquilegia* genus, their GC content were 33.01 (*matK*) and 45.10 (*rbcL*), respectively. It is possible that *matK,* a gene with low GC content and low GC preferences in the third position, may be prone to large sequence variation, indicating that in general genes with low GC content may play a more relevant role in clarifying the relationships among *Aquilegia* taxa.

Genes with lower GC preferences in the third codon position (such as *matK, ndhF*) usually are more ploymorphic than genes with higher GC content, indicating that the GC content may correlate with amino acid usage in chloroplast genes of *Aquilegia*. As some other groups also observed that GC content had a strong correlation with amino acid usage and codon usage (Wan et al., 2004; Tatarinova et al., 2010; Sablok et al., 2011), this conclusion may provide a general view of patterns in evolutionary studies in angiosperms.

In other land plants, such as *Oncidium gower ramsey* (Xu et al., 2011) and *Coffea arabica* (Nair et al., 2012), chloroplast genes has higher frequencies in NNT/NNA codons than in NNC/NNG codons. This phenomena also occurred in the *Aquilegia* genus. Zhou and his colleagues previously only showed that Valine had different preferred codons among chloroplast genes in six seed plants (Zhou et al., 2008), so their conclusion differed from the findings in this study. One possibility to explain this discrepancy is that

maybe they discarded the genes whose length is less than 300bp, and this would affect the proportion of codon usage. This was the case also in other studies, which only calculated codon usage of the *rbcL* gene (see, e.g., Liu and Xue, 2005). This study demonstrates that Valine may be a factor associated to the evolutionary divergence between eudicots and monocots, and detailed information on this amino acids needs more tests. The patterns of codon usage in chloroplast genes are association with GC content and this will help us to understand patterns of molecular evolution of *Aquilegia* in general. In contrast, highly conserved sequences means little variation among nucleotides and little phylogenetic information. Thus, these resulted in the poorly resolved relationships among taxa. Further studies of codon usage bias of mitochondrial, nuclear and chloroplast genes would help to clarify the relationship among these three genomes in *Aquilegia* genus.

4.1.2 Phylogenetic analyses in a highly conserved genus

A total of 34 taxa of *Aquilegia* which contained 28 species from Europe were sequenced, and 66 CDS and whole plastome (incomplete) datasets were used to construct the European clade relationships in this study. Although a distinct evolutionary relationships for all *Aquilegia* species was not clarified by either chloroplast CDS or whole plastomes, the Asian clade was well supported to be separate from the European clade. Additionally, several clusters in the European clade of this genus were supported by previous studies (Fior et al., 2013; Li et al., 2014). A recent study used only 21 non-coding regions of chloroplast genome to construct phylogenetic relationships among 84 *Aquilegia* taxa, but did not obtain obvious evolutionary relationships among European taxa (Fior et al., 2013). Taken together, this study illustrates that plastomes are sufficiently complex and therefore had higher capacity to capture phylogenetic information for closely clades than coding genes and partial non coding sequences (Parks et al., 2009). In contrast to the clear resolution at deeper branches, chloroplast data in this study resulted in poor resolution at some phylogenetic levels, such as the very recent and fast European radiation, which was characterized by a largely pectinate structure. Actually, the highly conserved nature and slow evolutionary rates of chloroplast genomes did not contain sufficient evolutionary information for a comparative study across species so closely related in the rapid radiation constituted by the European clade of the *Aquilegia* genus. Therefore, recently developed techniques based on sequencing of mitochondrial or nuclear DNA could be employed to further aid in the resolution of phylogenies at the genus level.

Nonetheless, the results obtained are relevant in furthering our understanding of the evolution of the *Aquilegia* genus, as they represent a significant improvement over previous studies. It is worth to note that the resolved clusters group largely based on

geography at different levels. The Asian taxa solidly group together, with the exception of *A. einseleana*, which is consistently clustering with them despite it is current mainly distributed in Italy. This incongruence can have different explanations. One possibility is that *A. einseleana* may have changed over time its distribution, either through range contraction from an original Asian location. Another possible scenario is that *A. einseleana* originated in the current distribution area as a hybrid between European taxa and a long-range dispersal event of an Asian taxon. As the chloroplasts are maternally inherited, this scenario is relatively unlikely, as it would imply a relatively large gene flow through seeds from Asia to Northern Italy. Unfortunately, with the data currently available, it is not possible to firmly conclude which of the above scenarios is at the base of the origin of *A. einseleana* at present. The use of nuclear markers will be required to elucidate the actual mechanism of this interesting case.

Other clusters identified in this study are more clearly supported by the distribution area of the species, indicating that the originated by vicariance. This is for instance the case of the Sardinian cluster formed by *A. barbaricina, A. nugorensis* and *A. nuragica* or by the one formed by *A. nigricans, A. reuterii* and *A. alpina*. The identification of these clades provides the exciting possibility to carry out in-depth ecological studies aimed to understand the ecological adaptations that took place in each clade and their contribution to differentiation and speciation by reproductive isolation.

4.2 Reconstructing the tribal relationships in the Brassicaceae family

4.2.1 Important positively selected genes

Based on positive selection analysis by Selecton, 33 genes were detected as under positive selection. The majority of these genes play fundamental roles in the most important and specific functions of the chloroplasts. It is known that many plastid proteins are part of the photosynthetic complexes, thus, some photosynthetic genes may be expected to be targeted by positive selection. ATP and NADPH molecules are necessary in photosynthesis for storing energy (Avenson et al., 2004). Consistently to the central relevance of bioenergetics in chloroplasts, some genes related to ATP and NADPH production were detected as under positive selection in this study, such as ATP synthase (3 genes), NADH dehydrogenase (6 genes) and ATP dependent protease. However, somehow more surprisingly, among the positively selected genes there are also some playing pivotal roles in chloroplast translation and transcription. The ribosomal proteins are fundamental building blocks for the ribosomal subunits involved in the cellular process of translation (Manuell et al., 2004). In this study, 6 and 5 genes for the large subunit and the small subunit ribosomal proteins of the chloroplast were detected, respectively. Besides, bacterial-type RNA polymerases are necessary for constructing

RNA chains using DNA genes as templates in chloroplast transcriptional process (Little and Hallick, 1988), while maturase is involved in splicing type II introns from RNA transcript. These enzymes are essential to life and are found in all organisms and many viruses. The *matK* and all the RNA polymerase genes (*rpoA, rpoB, rpoC1* and *rpoC2*) were detected as under positive selection in this study. It has been reported that the chloroplast *clpP* and *matK* genes are under positive selection in many plant such as *Oenothera, Lychnis chalcedonica,* and Cephalotaxaceae (Erixon and Oxelman, 2008; Chen and Xiao, 2010). The fact that, in addition to the major energy-related functions of the chloroplast, also transcription and translation are positively selected indicates that the expression levels of chloroplast-encoded genes may also have played and possibly still play a relevant role in adaptation processes in Brassicaceae.

A previous study has revealed that cytochrome genes (*petD* and *ccsA*) are detected as under positive selection in Brassicaceae species (Hu et al., 2015), and they were also detected in this study. Cytochromes are primarily responsible for the generation of ATP via electron transport in photosynthetic process, especially, cytochrome c for catalyzing redox reaction (Allen, 2003). The *psbD* gene is coding for D1 and D2 subunits of the photosystem II complex (Kiss et al., 2012), and is predicted under positive selection in this study. In contrast, the *psbA* undergo gene duplication in leptosporangiate and is detected under positive selection in ferns (Sen et al., 2012). However, such burst of high rate of evolution in Brassicaceae *psbA* was not observed, indicating that the encoded core protein complex in Brassicaceae (such as *psbB* and *psbC*) have a distinct evolutionary history as compared to that elucidated in basal plants.

Furthermore, other minor functional categories of chloroplast genes were also detected as under positive selection in this study. Most studies for positive selection analyses in conserved reading frame genes focused on the *ycf1* gene (Hu et al., 2015; Carbonell-Caballero et al., 2015; Yang et al., 2016). Although the *ycf1* gene was not detected in this study, the *ycf4* gene was under positive selection. The *ycf4* gene, encoding a thylakoid protein that plays a role in regulating photosystem I assembly, has undergone adaptive evolution in *Lathyrus* (Magee et al., 2010). The *cemA* gene was named *ycf10* before the cemA protein was identified in the inner envelope membrane of chloroplasts. The cemA protein induces light sensitivity, but does not affect cell viability or photosynthetic reactions in plants. Zhong and his colleagues detected the *cemA* gene as under positive selection in *Poaceae* (Zhong et al., 2009). The pseudogene *accD,* first defined in *A. syriaca* (Straub et al., 2012), encodes a subunit of the acetyl-CoA carboxylase in dicots, which regulates fatty-acid biosynthesis in the plastid (Lam et al., 2015). In addition, this gene has been functionally relocated to nucleus in Campanulaceae (Rousseau-Gueutin et al., 2013). The identification of positive selection in *accD* indicates that this gene may be repeatedly involved in the adaptation to specific ecologic niches during the radiation of plants.

Owing to its use as one of the most common phylogenetic markers in plants (Savolainen et al., 2000; Wink and Mohamed, 2003; Janssen and Bremer, 2004), the plastome gene encoding for ribulose-1, 5-bisphospate carboxylase/oxygenase large subunit, *rbcL*, has been object of extensive evolutionary analyses. It encodes the large subunit of Rubisco, which is responsible for the organization of atmospheric CO_2. Despite the enormous relevance of this enzyme, Rubisco did not evolve the ability to specifically recognize CO_2 from O_2, the major subproduct of the water-splitting process associated to photosynthesis. This lack of specificity results in an energy-wasting by light reaction of photosynthesis, called photorespiration (Kapralov and Filatov, 2007). Sen et al. studied evolution of the *rbcL* gene in three gymnosperm families, and found that seven sites were positively selected in these families (Sen et al., 2011). Christin and his colleagues demonstrated that *rbcL* was detected as under positive selection in C4 lineage plants through 338 monocot species (Christin et al., 2008). Barrett and Freudenstein illustrated that *rbcL* was under positive selection in 31 accessions of 8 species of *Corallorhiza*, and ω was influenced by stop codons and frameshifts (Barrett and Freudenstein, 2008). Su and Wang observed that seven sites of *rbcL* were detected under positive in Polypodiaceae (Su and Wang, 2008). Hao et al. studies 393 species from 11 gymonsperm groups and found that the presence of positive selection in *rbcL* was observed in all of the gymnosperm groups (Hao et al., 2010). Additionallly, Kapralov and Filatov searched more than 3000 species representing all lineages of green plants and some lineages of other phototrophs for adaptive evolution in *rbcL* sequences, and found that adaptive evolution existed in *rbcL* of most analyzed plants (Kapralov and Filatov, 2007). Taken together, these studies indicated that patterns of molecular evolution in the *rbcL* gene in different families or lineages may be similar.

Through Selecton program, 9 sites (281A, 326I, 445I, 472I, 477K, 480E, 481D, 482E and 483E) of *rbcL* were significantly positively selected in this analysis. Previous studies reported in Brassicaceae that amino acids 281, 472 and 477 were identified among *rbcL* residues evolving under positive selection (Iida et al., 2009; Hu et al., 2015), and these three sites were also observed in this study. The residue (326) is in close proximity to the fourth among the most often positively selected Rubisco residues in plants (amino acid 328), which has been associated to adaptive variation of Rubisco active site possibly by modifying the position of 327. Thus, in Brassicaceae residue 326 may affect Rubisco discrimination between CO_2 and O_2 fixation, analogously to what suggested for residue 328 in several other plant groups (Wang et al., 2011). This situation also occurred in 445 residue, as it could be supposed to have analogous function for residue 449, which involved in the regulation of the conformational changes promoted by the oxidative modification delaying process associated with the catabolism of enzymes in vivo (Marin-Navarro and Moreno, 2006). The other four residues (480-483) have not been detected as positively selection sites in other studies. It is possible that they may be caused by

interactions among these sites, like coevolution among other active residues in *rbcL* (Sen et al., 2011). As positive selection could be detected by identifying cases where $\omega > 1$, thus, it is very important to account for variation in selective pressure among sites if one hopes to detect positive selection affecting only a few amino acid residues or influencing the evolutionary rates. Therefore, further tests will be required for the complete elucidation of the patterns of selective pressure acting on these genes in the future, such as branch level analysis. Combination of different programs could provide more precise results. For instance, the MEME program would identify both episodic and pervasive positive selection at the site level (Murrell et al., 2012b).

4.2.2 The classification of the Brassicaceae family

The phylogenetic relationships on 71 CDS from 95 Brassicaceae species over 33 tribes were reconstructed and three major lineages were defined supporting by previous studies (Beilstein et al., 2006). Besides the major three lineages, the tribe Thlaspideae was problematic, as it was not monophyletic. The four sampled species were clustered into one lineage, but only *Peltaria angustifolia* was inferred to take a position close to the tribe Eutremeae (Al-Shehbaz et al. 2006), while the other three species were close to Biscutelleae (Zunk et al., 1999). A similar situation was observed in other tribes, such as the Camelineae tribe (Bailey et al., 2006; Couvreur et al., 2010) and the Anchonieae tribe (Warwick et al., 2007). Besides, a previous study showed that Thlaspideae was separated into two parts by Eutremeae (Beilstein et al., 2008). The reason causing this separation is most likely due to the fact that insufficient phylogenetic information was available and thus difficult relationships among tribes of Brassicaceae could not be clarified. Another problem is the tribe Anastaticeae, which was surrounded by the tribes belonging to expanded lineage II in this phylogeny while the position of this tribe in previous study analyzed by whole-genome duplications was in lineage III (Franzke et al., 2011). But in other phylogeny summaries, this tribe was also located inside the range of expanded lineage II and also closely related to the Iberideae tribe (Beilstein et al., 2006; Huang et al., 2015). Thus, the support from both chloroplast and nuclear data suggests the possible reclassification of the taxonomic position of this tribe.

The results of this study supported the three previously recognized lineages that were first discovered with chloroplast *ndhF* sequence (Beilstein et al., 2006). However, the final phylogenetic tree in this study still has some conflicts with previous studies, such as some locations of tribes inside of lineages. For instance, 1) the tribe Alysseae, which belonged to expanded lineage II and split Anastaticeae from other linage III tribes, was closely related to subtree of Biscutelleae and Cochlearieae (Huang et al., 2015). But it was observed between Biscutelleae and Cochlearieae in this study. 2) The tribe Lepidieae (lineage I), which took the evolutionary position between Cardamineae and Erysimeae in

Kagale's phylogeny (Kagale et al., 2014), was identified as outgroup to both of these tribes in this study. 3) The relationships of Turritititideae, Erysimeae and Alyssopsideae tribes. Like in this study, Erysimeae and Alyssopsideae were sister to each other, and they were distantly related to Turritititideae. By contrast in Huang's results, Turritititideae and Alyssopsideae were sister to each other (Huang et al., 2015). One has to notice that these results were from chloroplast genes, while other studies were from nuclear genes and transcriptome data, indicating that different sequence data would have inconsistent topologies and relationships. Reassuringly, however, this study supported the results of previous chloroplast analyses in Brassicaceae (Beilstein et al., 2006). However, analysis of only chloroplast genes could not clarify the real relationships in Brassicaceae due to their maternal inheritance. In light of the late origin of Brassicaceae, also the wide hybridization and polyploidy, this means that an approach with only sequences from the chloroplast genome may not be sufficient to solve the complete phylogeny of this family, especially for the most problematic clades. For a complete understanding of the evolutionary history of the Brassicaceae family, the combined analysis with the nuclear genome should be a promising approach.

4.3 Patterns of chloroplast evolution across families in angiosperms

4.3.1 Pattern of gene wide level selection

As indicated in the Material and Methods section, the results of the single BUSTED analyses were corrected by applying a false discovery rate multiple test correction to account for the statistical family of hypotheses that have to be contemporary assessed across taxonomic families and genes to infer recurrent patterns of selection. The FDR concept was formally described by Benjamini and Hochberg in 1995 (Benjamini and Hochberg, 1995) as a way of conceptualizing the rate of type I errors in null hypothesis testing when conducting multiple comparisons. Application of FDR to this study, in other words, becomes necessary if one wants to compare with each other all families and genes by controlling the rate of type I errors (false positive). In this case one has to consider the result of each BUSTED analysis as a single statistical test of the null hypothesis that loosely can be verbally formulated as: "Is positive selection acting randomly across chloroplast genes in all the 11 families considered as independent replicates of plant species evolution?". This approach is clearly an approximation based on the assumption that all genes as well as the species in each family have the same divergence, so that the information content of each alignment is the same to that of all others. From literature, it is known that chloroplast genes evolve at different speeds and from a cursory analysis of the alignments it is also clear that not all of them contain the same number of species and that the phylogenetic distance and distribution of taxa varies from family to family.

Despite these limitations, the FDR correction is expected to be conservative, as it should detect true positively selected genes at the price of lower statistical power, given the high number of false negative results which it could allow. In other words, alignments with low divergence and/or few and closely related taxa should fail to be identified among the positives even though effectively under positive selection, while more informative alignments which are under positive selection should be successfully identified even by applying the FDR correction.

The *rbcL* gene was the only major outlier of the number of BUSTED hits per family after FDR correction by p-values. Botanists have very often chosen *rbcL* for phylogenetic reconstructions in thousands of plant species due to its very high conservation, presence in high copy number and ease of amplification from a wide taxonomic range of species with a few highly conserved primers (Källersjö et al., 1998; Soltis et al., 2000; Tamura et al., 2004). However, the high likelihood of undergoing positive selection detected in this and former studies (Kapralov and Filatov, 2007) requires further consideration, as it could potentially affect phylogenetic reconstruction. Noteworthy, Kapralov and Filatov analyzed it in over 3000 species of green plants and other phototrophs, and found that *rbcL* is under positive selection in most land plants, but not in algae and cyanobacteria (Kapralov and Filatov, 2007). The observation that all 6 outliers found in the distribution of alignment polymorphic levels are among the genes with multiple BUSTED hits warrants some caution in the interpretation of the results as it indicates that more polymorphic loci tend to be identified more readily than less polymorphic ones. As noted above, this suggest that, at least in part, the identification of positive selection signatures by BUSTED may lack power in less polymorphic genes, making the test over-conservative. In other words, it is possible that the analysis carried out may have missed additional genes whose polymorphism levels were too low for reliable positive selection detection by BUSTED. From this point of view, the genes at the left tail of the distribution of the polymorphism level (Figure 3.7) may constitute relevant candidates for functional studies even though they were identified just in single families (e.g. *psbF, rps7, rps16, rps18* and several other subunits of the ribosomes). In most of the families analyzed, evidence of 2 or more genes were observed as under positive selection, except in the Apiaceae family where no gene putatively under selection could be identified.

4.3.2 Pattern of site level selection

The application of multiple test correction to the results of the MEME analyses is overly conservative, for the same reasons noted above. It is therefore likely that additional sites discarded by this procedure may be worth further investigation, as one can expect a relatively large number of false positives. While for the majority of plastome-encoded proteins no three-dimensional structure is available, it is currently difficult to propose a

putative structure and/or functional hypothesis which could explain the selective patterns detected for the *ccsA* and *rpoC2* genes. Based on the crystal structure, the position 174 in the *psaA* alignment corresponds to the position 201 of the psaA peptide in the photosystem I (PSI) PDB used as reference (4xk8.1). The position is occupied by small amino acid residues and it is located in the middle of a membrane spanning alpha helix. The serine side chain most commonly present in the Rosaceae alignment is located in proximity and involved in the binding of one chlorophyll a and one beta-carotene molecule. Given the high number of molecules of chlorophyll and beta-carotene bound to the PSI complex, it is currently difficult to predict any special role that could explain the selective patterns observed for this residue. On the other hand, RbcL position 247 is responsible for the only inter-subunit disulphide bridge experimentally validated till now. While mutation of this residue does not seem to affect the catalytic activity of the enzyme (Marcus et al., 2003), Cys247 is one of the three cysteine residues undergoing both nitrosylation and glutathionylation, two forms of post translational modification thought to be involved in the redox regulation of Rubisco activity and/or stability (Sudhani, 2012). Taken together, these observations point to a possible lineage-specific role of Cys247 in regulation of the aggregation status of Rubisco, but functional studies will be required to experimentally test this hypothesis.

4.3.3 Pattern of branch specific selection

Besides gene-wide and site level selection, 46 branches were identified under positive selection after FDR correction. Outlier detection based on IQR test identified a total of 6 (4 minor and 2 major) outlier genes with respect to the overall distribution of times each gene was hit as positively selected in a specific branch. The majority of the outliers identified in BUSTED analyses (*ccsA, rpoB* and *ycf1*) were consistently identified as outliers also in aBSREL analyses, indicating the congruence of branch-specific and gene-wide patterns of positive selection. Three additional genes (*rpoC1, ycf2* and *ndhF*) were detected as outliers only in aBSREL analyses, indicating that episodic positive selection may be the major evolutionary pattern affecting these genes. The major outlier of BUSTED analyses, *rbcL*, was not detected as outlier in aBSREL, suggesting that the pattern of gene-wide selection is not specifically associated to any particular branch or set of branches of the gene phylogeny, but is the result of a relatively uniform selection acting evenly in time across lineages. In addition, the *rpoC2* gene of Caryophyllaceae and the *ccsA* gene of Fabaceae were also supported by MEME.

Also noteworthy, the majority of the branches detected as under positive selection were terminal in the tree, i.e. they can be traced to most recent evolutionary history of extant species. The ability to detect positive selection in an interspecific context is a function of the ecological replacements among species with analogous niches. Given the

conservation of species ecological adaptations and the low rate of colonization of novel habitats (Soltis et al., 2004), it is possible that the relatively high number of recent branches under positive selection may be the result of a lower ability to detect ancient patterns of positive selection, as they would be obscured by taxonomical replacement favoring taxa bearing positively selected variants of the gene(s). Alternatively, the dynamic nature over time of flora assemblages due to recurrent paleoclimatic changes may allow to trace one of the most recent patterns of selection associated to the last contraction and expansion of range in the local flora. As also demographic factors like effective population size can affect the species-specific patterns of molecular evolution, it is possible that range contraction and expansion could have been contributing at different levels to the generation of the observed selection patterns.

4.3.4 Relaxed selection

It is commonly known that natural populations often undergo the weakening or complete removal of a source of selection that has been important in the maintenance of one or more traits. Biologists refer to these situations as 'relaxed selection', and explore the effects of such changes on traits in their ecological contexts (Lahti et al., 2009). The hypothesis testing framework RELAX (Wertheim et al., 2015) which is a package of the HyPhy program was used to test for potentially relaxed selection of protein coding genes across families. RELAX calculates a selection intensity parameter, k, taking into account that relaxation would have opposite effects on sites subjected to purifying selection ($\omega < 1$) and sites subjected to positive selection ($\omega > 1$), and relaxation would move ω toward 1 for both categories. The likely ancestral state of the internal nodes of the phylogeny of each species was not reconstructed to avoid propagation of modelling uncertainty to the assessment of the selective pattern differences among groups. Thus, only the analysis of extant species was carried out in this study, treating all internal nodes as missing data.

A previous study in ants reported that fast-evolving genes were preferentially recruited into caste-biased gene expression, especially the genes under relaxed selective constraints (Hunt et al., 2011). In general, genes under strong positive selection would have a Ka/Ks ratio (ω) greater than 1. In the case of BUSTED, all of the four genes identified had fast substitution rates. The results are interesting from different points of view. First of all, they indicate that relaxed selection seems to play a more relevant, but not exclusive, role than intensified, directional selection in the adaptation to high elevation. Large functional variation in photosynthetic performance associated to the transition from autotrophy to heterotrophy in parasitic plants has been associated to prolonged and progressively increasing relaxed selection, ultimately leading to gene and functional loss (Wicke et al. 2014; Wicke et al., 2016). The number and intensity of events of relaxed selection detected in this study are much smaller than those reported in the above case, but it is

nonetheless very significant that such signatures could be detected in association to elevation preferences. The mechanistic underpinning of the observed patterns of selection are currently unknown. It is, however, noteworthy that a ribosomal subunit was recently implicated in the adaptation to the parasitic life style in Cytinaceae, suggesting that translational efficiency may be involved also in the intensified selection detected in the *rps18* gene (Roquet et al., 2016). Noteworthy, in the same family (Fabaceae) the *rpoC1* gene underwent relaxed selection in association with elevation, indicating that the pattern observed for this gene may be an indirect consequence of an increased translational efficiency in chloroplast, thus reducing the selective pressure on transcription. By contrast both of the genes detected by RELAX in Asteraceae experienced relaxed selection. The very low K value detected for *accD* may be an artifact due to the high variability of this gene, but it indicates that a very low selecting pressure has been acting on this gene in the Asteraceae. The moderate relaxed selection detected in the *rbcL* gene may instead be the result of a decrease in the oxygenase activity of Rubisco with elevation, as both a temperature decrease and the differential solubility variation of CO_2 and O_2 would theoretically favour the carboxylase activity. The reason why this should be specifically related to the Asteraceae family may reside in its very high dispersal ability, which may lead to a fast replacement rate among high elevation species and the persistence of the selective pattern on the gene.

References

Aghaei, R., Alvaninejad, S., Zolfaghari, R., Reza, M., & Gharehlar, R. (2013). Flora, life form and geographical distribution of plants in West-South forests of Iran (Case study: Vezg, Yasouj). *International Journal Farming Allied Science*, *2*, 1046-1053.

Akaike, H. (1998). Information theory and an extension of the maximum likelihood principle. In *Selected Papers of Hirotugu Akaike* (pp. 199-213). Springer New York.

Allen, J. F. (2003). Cyclic, pseudocyclic and noncyclic photophosphorylation: new links in the chain. *Trends in Plant Science*, *8*(1), 15-19.

Al-Shehbaz, I. A., & O'Kane Jr, S. L. (2002). *Lesquerella* is united with *Physaria* (Brassicaceae). *Novon*, *12*(3), 319-329.

Al-Shehbaz, I. A., Beilstein, M. A., & Kellogg, E. A. (2006). Systematics and phylogeny of the Brassicaceae (Cruciferae): an overview. *Plant Systematics and Evolution*, *259*(2-4), 89-120.

Ančev, M. (2007). Catalogue of the family Brassicaceae (Cruciferae) in the flora of Bulgaria. *Phytologia Balcanica*, *13*, 153-178.

Andrews, S. (2010). FastQC: a quality control tool for high throughput sequence data. 175-176. http://www.bioinformatics.babraham.ac.uk/projects/fastqc/

Anisimova, M., & Gascuel, O. (2006). Approximate likelihood-ratio test for branches: A fast, accurate, and powerful alternative. *Systematic Biology*, *55*(4), 539-552.

Arrigoni, R., Vacherie, B., Benzoni, F., Stefani, F., Karsenti, E., Jaillon, O., Not, F., Nunes, F., Payri, C., Wincher, P., & Barbe, V. (2017). A new sequence data set of SSU rRNA gene for Scleractinia and its phylogenetic and ecological applications. *Molecular Ecology Resources*. doi:10.1111/1755-0998.12640

Assefa, S., Keane, T. M., Otto, T. D., Newbold, C., & Berriman, M. (2009). ABACAS: algorithm-based automatic contiguation of assembled sequences. *Bioinformatics*, *25*(15), 1968-1969.

Avenson, T. J., Cruz, J. A., & Kramer, D. M. (2004). Modulation of energy-dependent quenching of excitons in antennae of higher plants. *Proceedings of the National Academy of Sciences of the United States of America*, *101*(15), 5530-5535.

Bailey, C. D., Al-Shehbaz, I. A., & Rajanikanth, G. (2007). Generic limits in tribe Halimolobeae and description of the new genus *Exhalimolobos* (Brassicaceae). *Systematic Botany*, *32*(1), 140-156.

Bailey, C. D., Koch, M. A., Mayer, M., Mummenhoff, K., O'kane, S. L., Warwick, S. I., Windham, M. D., & Al-Shehbaz, I. A. (2006). Toward a global phylogeny of the Brassicaceae. *Molecular Biology and Evolution*, *23*(11), 2142-2160.

Ballerini, E. S., & Kramer, E. M. (2011). Environmental and molecular analysis of the floral transition in the lower eudicot *Aquilegia formosa*. *EvoDevo*, *2*(1), 1-20.

Barow, M., & Meister, A. (2002). Lack of correlation between AT frequency and genome size in higher plants and the effect of nonrandomness of base sequences on dye binding. *Cytometry*, *47*(1), 1-7.

Barrett, C. F., & Freudenstein, J. V. (2008). Molecular evolution of *rbcL* in the mycoheterotrophic coralroot orchids (Corallorhiza Gagnebin, Orchidaceae). *Molecular Phylogenetics and Evolution*, *47*(2), 665-679.

Bastida, J. M., Alcántara, J. M., Rey, P. J., Vargas, P., & Herrera, C. M. (2010). Extended phylogeny of *Aquilegia*: the biogeographical and ecological patterns of two simultaneous but contrasting radiations. *Plant Systematics and Evolution*, *284*(3-4), 171-185.

Becker, M., Gruenheit, N., Deusch, O., Voelckel, C., & Lockhart, P.J. (2013). Nunatak survival in the Central Southern Alps of New Zealand. Unpublished.

Beilstein, M. A., Al-Shehbaz, I. A., & Kellogg, E. A. (2006). Brassicaceae phylogeny and trichome evolution. *American Journal of Botany*, *93*(4), 607-619.

Beilstein, M. A., Al-Shehbaz, I. A., Mathews, S., & Kellogg, E. A. (2008). Brassicaceae phylogeny inferred from phytochrome A and *ndhF* sequence data: tribes and trichomes revisited. *American Journal of Botany*, *95*(10), 1307-1327.

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B (Methodological)*, 289-300.

Bleeker, W., & Matthies, A. (2005). Hybrid zones between invasive *Rorippa austriaca* and native *R. sylvestris* (Brassicaceae) in Germany: ploidy levels and patterns of fitness in the field. *Heredity*, *94*(6), 664-670.

Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics (Oxford, England)*, *30*(15), 2114-2120.

Bremer, B., Bremer, K., Chase, M., Fay, M., Reveal, J., Soltis, D., Soltis, P., & Stevens, P. (2009). An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG III. *Botanical Journal of the Linnean Society*, *161*(2), 105-121.

Bremer, B., Bremer, K., Heidari, N., Erixon, P., Olmstead, R. G., Anderberg, A. A., Källersjö, M., & Barkhordarian, E. (2002). Phylogenetics of asterids based on 3 coding and 3 non-coding chloroplast DNA markers and the utility of non-coding DNA at higher taxonomic levels. *Molecular Phylogenetics and Evolution*, *24*(2), 274-301.

Bungard, R. A. (2004). Photosynthetic evolution in parasitic plants: insight from the chloroplast genome. *Bioessays*, *26*(3), 235-247.

Busset, J., Cabau, C., Meslin, C., & Pascal, G. (2011). PhyleasProg: a user-oriented web server for wide evolutionary analyses. *Nucleic Acids Research*, *1*, 7.

Caetano-Anollés, G. (2005). Evolution of genome size in the grasses. *Crop Science*, *45*(5), 1809-1816.

Cai, Z., Guisinger, M., Kim, H. G., Ruck, E., Blazier, J. C., McMurtry, V., Kuehl, J. V., Boore, J., & Jansen, R. K. (2008). Extensive reorganization of the plastid genome of *Trifolium subterraneum* (Fabaceae) is associated with numerous repeated sequences and novel DNA insertions. *Journal of Molecular Evolution*, *67*(6), 696-704.

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., & Madden, T. L. (2009). BLAST+: architecture and applications. *BMC Bioinformatics*, *10*(1), 1-9.

Carbonell-Caballero, J., Alonso, R., Ibañez, V., Terol, J., Talon, M., & Dopazo, J. (2015). A phylogenetic analysis of 34 chloroplast genomes elucidates the relationships between wild and domestic species within the genus *Citrus*. *Molecular Biology and Evolution*, *32*(8), 2015-2035.

Castresana, J. (2000). Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Molecular Biology and Evolution*, *17*(4), 540-552.

Cavalier-Smith, T. (2000). Membrane heredity and early chloroplast evolution. *Trends in Plant Science*, *5*(4), 174-182.

Chan, C. X., Gross, J., Yoon, H. S., & Bhattacharya, D. (2011). Plastid origin and evolution: new models provide insights into old problems. *Plant Physiology*, *155*(4), 1552-1560.

Charlesworth, D., Liu, F. L., & Zhang, L. (1998). The evolution of the alcohol dehydrogenase gene family by loss of introns in plants of the genus *Leavenworthia* (Brassicaceae). *Molecular Biology and Evolution*, *15*(5), 552-559.

Chaw, S. M., Zharkikh, A., Sung, H. M., Lau, T. C., & Li, W. H. (1997). Molecular phylogeny of extant gymnosperms and seed plant evolution: analysis of nuclear 18S rRNA sequences. *Molecular Biology and Evolution*, *14*(1), 56-68.

Chen, S. L., & Xiao, P. G. (2010). Molecular evolution and positive Darwinian selection of the chloroplast maturase matK. *Journal of Plant Research*, *123*(2), 241-247.

Christenhusz, M. J., & Byng, J. W. (2016). The number of known plants species in the world and its annual increase. *Phytotaxa*, *261*(3), 201-217.

Christin, P. A., Salamin, N., Muasya, A. M., Roalson, E. H., Russier, F., & Besnard, G. (2008). Evolutionary switch and genetic convergence on *rbcL* following the evolution of C4 photosynthesis. *Molecular Biology and Evolution*, *25*(11), 2361-2368.

Conti, F., Abbate, G., Alessandrini, A., & Blasi, C. (2005). An Annotated Checklist of the Italian Vascular Flora. Ministero dell'Ambiente e della Tutela del Territorio, Direzione per la Protezione della Natura. Dipartimento di Biologia Vegetale Università "La Sapienza", Ed. Palombi, Rome, Italy.

Couvreur, T. L., Franzke, A., Al-Shehbaz, I. A., Bakker, F. T., Koch, M. A., & Mummenhoff, K. (2010). Molecular phylogenetics, temporal diversification, and principles of evolution in the mustard family (Brassicaceae). *Molecular Biology and Evolution*, *27*(1), 55-71.

Creevey, C. J., & McInerney, J. O. (2005). Clann: investigating phylogenetic information through supertree analyses. *Bioinformatics*, *21*(3), 390-392.

Cronn, R., Liston, A., Parks, M., Gernandt, D. S., Shen, R., & Mockler, T. (2008). Multiplex sequencing of plant chloroplast genomes using Solexa sequencing-by-synthesis technology. *Nucleic Acids Research*, *1*, 11.

Danecek, P., Nellåker, C., McIntyre, R. E., Buendia-Buendia, J. E., Bumpstead, S., Ponting, C. P., Flint, J., Durbin, R., Keane, T. M., & Adams, D. J. (2012). High levels of RNA-editing site conservation amongst 15 laboratory mouse strains. *Genome Biology*, *13*(4), 1-12.

Daniell, H., Lin, C. S., Yu, M., & Chang, W. J. (2016). Chloroplast genomes: diversity, evolution, and applications in genetic engineering. *Genome Biology*, *17*(1), 1-29.

Darling, A. C., Mau, B., Blattner, F. R., & Perna, N. T. (2004). Mauve: multiple alignment of conserved genomic sequence with rearrangements. *Genome Research*, *14*(7), 1394-1403.

Darriba, D., Taboada, G. L., Doallo, R., & Posada, D. (2011). ProtTest 3: fast selection of best-fit models of protein evolution. *Bioinformatics*, *27*(8), 1164-1165.

Darriba, D., Taboada, G. L., Doallo, R., & Posada, D. (2012). jModelTest 2: more models, new heuristics and parallel computing. *Nature Methods*, *9*(8), 772-772.

de Vos, A. C., Broekman, R., de Almeida Guerra, C. C., van Rijsselberghe, M., & Rozema, J. (2013). Developing and testing new halophyte crops: A case study of salt tolerance of two species of the Brassicaceae, *Diplotaxis tenuifolia* and *Cochlearia officinalis*. *Environmental and Experimental Botany*, *92*, 154-164.

Delsuc, F., Brinkmann, H., & Philippe, H. (2005). Phylogenomics and the reconstruction of the tree of life. *Nature Reviews Genetics*, *6*(5), 361-375.

Dilcher, D. L., & Crane, P. R. (1984). Archaeanthus: an early angiosperm from the cenomanian of the western interior of North America. *Annals of the Missouri Botanical Garden*, *71*(2), 351-383.

Dorn, R. D. (2003). A new species of *Boechera* (Brassicaceae) from Utah and Colorado. *Brittonia*, *55*(1), 1-3.

Doyle, J. A. (2006). Seed ferns and the origin of angiosperms. *The Journal of the Torrey Botanical Society*, *133*(1), 169-209.

Drory, O., Frolow, F., & Nelson, N. (2004). Crystal structure of yeast V-ATPase subunit C reveals its stator function. *EMBO Reports*, *5*(12), 1148-1152.

Drummond, D. A., & Wilke, C. O. (2008). Mistranslation-induced protein misfolding as a dominant constraint on coding-sequence evolution. *Cell*, *134*(2), 341-352.

Duff, R. J., & Nickrent, D. L. (1999). Phylogenetic relationships of land plants using mitochondrial small-subunit rDNA sequences. *American Journal of Botany*, *86*(3), 372-386.

Erixon, P., & Oxelman, B. (2008). Whole-gene positive selection, elevated synonymous substitution rates, duplication, and indel evolution of the chloroplast *clpP1* gene. *PLoS One*, *3*(1), e1386.

Ferrarini, M., Moretto, M., Ward, J. A., Šurbanovski, N., Stevanović, V., Giongo, L., Viola, R., Cavalieri, D., Velasco, R., Cestaro, A., & Sargent, D. J. (2013). An evaluation of the PacBio RS platform for sequencing and *de novo* assembly of a chloroplast genome. *BMC Genomics*, *14*(1), 670.

Fior, S., Li, M., Oxelman, B., Viola, R., Hodges, S. A., Ometto, L., & Varotto, C. (2013). Spatiotemporal reconstruction of the *Aquilegia* rapid radiation through next-generation sequencing of rapidly evolving cpDNA regions. *New Phytologist*, *198*(2), 579-592.

Fonderflick, J., Lepart, J., Caplat, P., Debussche, M., & Marty, P. (2010). Managing agricultural change for biodiversity conservation in a Mediterranean upland. *Biological Conservation*, *143*(3), 737-746.

Franzke, A., German, D., Al-Shehbaz, I. A., & Mummenhoff, K. (2009). Arabidopsis family ties: molecular phylogeny and age estimates in Brassicaceae. *Taxon*, *58*(2), 425-437.

Franzke, A., Lysak, M. A., Al-Shehbaz, I. A., Koch, M. A., & Mummenhoff, K. (2011). Cabbage family affairs: the evolutionary history of Brassicaceae. *Trends in Plant Science*, *16*(2), 108-116.

Fulton, M., & Hodges, S. A. (1999). Floral isolation between *Aquilegia formosa* and *Aquilegia pubescens*. *Proceedings of the Royal Society of London B: Biological Sciences*, *266*(1435), 2247-2252.

Garnock-Jones, P. J. (1981). Checklist of dicotyledons naturalised in New Zealand 9. Ceratophyllaceae, Nymphaeaceae, Ranunculaceae, and Violaceae. *New Zealand Journal of Botany*, *19*(2), 157-160.

Gaut, B., Yang, L., Takuno, S., & Eguiarte, L. E. (2011). The patterns and causes of variation in plant nucleotide substitution rates. *Annual Review of Ecology, Evolution, and Systematics*, *42*, 245-266.

Gil, M., Zanetti, M. S., Zoller, S., & Anisimova, M. (2013). CodonPhyML: fast maximum likelihood phylogeny estimation under codon substitution models. *Molecular Biology and Evolution*, *30*(6), 1270-1280.

Gilardelli, F., Gentili, R., Prosser, F., Bonomi, C., Varotto, C., & Sgorbati, S. (2013). Ecological and biodiversity gradients across alpine dry grassland habitats: implications for an endangered species. *Nordic Journal of Botany*, *31*(2), 225-238.

Gilson, P. R., Su, V., Slamovits, C. H., Reith, M. E., Keeling, P. J., & McFadden, G. I. (2006). Complete nucleotide sequence of the chlorarachniophyte nucleomorph: nature's smallest nucleus. *Proceedings of the National Academy of Sciences*, *103*(25), 9566-9571.

Glöckner, G., Rosenthal, A., & Valentin, K. (2000). The structure and gene repertoire of an ancient red algal plastid genome. *Journal of Molecular Evolution*, *51*(4), 382-390.

Goldman, N., & Yang, Z. (1994). A codon-based model of nucleotide substitution for protein-coding DNA sequences. *Molecular Biology and Evolution*, *11*(5), 725-736.

Gordon, A., Hannon, G. J. (2012). "FASTX-Toolkit", FASTQ/A short-reads pre-processing tools (unpublished). http://hannonlab.cshl.edu/fastx_toolkit/

Gould, S. J. (1998). Gulliver's further travels: the necessity and difficulty of a hierarchical theory of selection. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, *353*(1366), 307-314.

Gratani, L., Catoni, R., & Varone, L. (2011). Photosynthetic and leaf respiration activity of *Malcolmia littorea* (L.) R. Br. in response to air temperature. *Photosynthetica*, *49*(1), 65-74.

Graur, D., & Li, W. H. (2000). Fundamentals of molecular evolution. *Sunderland, MA Sinauer Associates,* 209–210.

Guacchio, E. D. (2009). *Aquilegia vulgaris* var. *speluncarum Lacaita* (Ranunculaceae): an enigmatic columbine from the Campanian Apennines, S Italy. *Willdenowia*, *39*(1), 63-68.

Guilloux, A., & Jestin, J. L. (2012). The genetic code and its optimization for kinetic energy conservation in polypeptide chains. *Biosystems*, *109*(2), 141-144.

Guimarães, M. D. A., Gustin, J. L., & Salt, D. E. (2009). Reciprocal grafting separates the roles of the root and shoot in zinc hyperaccumulation in *Thlaspi caerulescens*. *New Phytologist*, *184*(2), 323-329.

Guindon, S., Dufayard, J. F., Hordijk, W., Lefort, V., & Gascuel, O. (2009). PhyML: fast and accurate phylogeny reconstruction by maximum likelihood. *Infection Genetics and Evolution, 9*(3), 384-385.

Hahn, M. W. (2007). Toward a selection theory of molecular evolution. *Evolution*, *62*(2), 255-265.

Hall, A. E., Fiebig, A., & Preuss, D. (2002b). Beyond the Arabidopsis genome: opportunities for comparative genomics. *Plant Physiology*, *129*(4), 1439-1447.

Hall, J. C., Iltis, H. H., & Sytsma, K. J. (2004). Molecular phylogenetics of core Brassicales, placement of orphan genera *Emblingia*, *Forchhammeria*, *Tirania*, and character evolution. *Systematic Botany*, *29*(3), 654-669.

Hall, J. C., Sytsma, K. J., & Iltis, H. H. (2002a). Phylogeny of Capparaceae and Brassicaceae based on chloroplast sequence data. *American Journal of Botany*, *89*(11), 1826-1842.

Hall, T. A. (1999). BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symposium Series,41*(41), 95-98.

Hallick, R. B., Drager, R. Hong, L., G., Favreau, M. R., Monfort, A., Orsat, B., Spielmann, A., & Stutz, E. (1993). Complete sequence of *Euglena gracilis* chloroplast DNA. *Nucleic Acids Research*, *21*(15), 3537-3544.

Hansen, A. K., Escobar, L. K., Gilbert, L. E., & Jansen, R. K. (2007). Paternal, maternal, and biparental inheritance of the chloroplast genome in *Passiflora* (Passifloraceae): implications for phylogenetic studies. *American Journal of Botany*, *94*(1), 42-46.

Hao, D. C., Mu, J., & Xiao, P. G. (2010). Molecular evolution and positive Darwinian selection of the gymnosperm photosynthetic Rubisco enzyme. *Botanical Studies*, *51*(4), 491-510.

Hilu, K. W., Borsch, T., Müller, K., Soltis, D. E., Soltis, P. S., Savolainen, V., Chase, M. W., Powell, M. P., Alice, L. A., Evans, R., Sauquet, H., Neinhuis, C., Slotta, T. A. B., Rohwer, J. G., Campbell, C. S., & Chatrou, L. W. (2003). Angiosperm phylogeny based on *matK* sequence information. *American Journal of Botany*, *90*(12), 1758-1776.

Hiratsuka, J., Shimada, H., Whittier, R., Ishibashi, T., Sakamoto, M., Mori, M., Kondo, C., Honji, Y., Sun, C. R., Meng, B. Y., Li, Y. Q., Kanno, A., Nishizawa, Y., Hirai, A., Shinozaki, K. & Sugiura,M. (1989). The complete sequence of the rice (*Oryza sativa*) chloroplast genome: intermolecular recombination between distinct tRNA genes accounts for a major plastid DNA inversion during the evolution of the cereals. *Molecular and General Genetics*, *217*(2-3), 185-194.

Hoch, B., Maier, R. M., Appel, K., Igloi, G. L., & Kössel, H. (1991). Editing of a chloroplast mRNA by creation of an initiation codon. *Nature*, *353*, 178-180.

Hodges, S. A. (1997). Floral nectar spurs and diversification. *International Journal of Plant Sciences*, *158*(S6), S81-S88.

Hodges, S. A., & Derieg, N. J. (2009). Adaptive radiations: from field to genomic studies. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(Suppl. 1), 9947-9954.

Hodges, S. A., Fulton, M., Yang, J. Y., & Whittall, J. B. (2004). Verne Grant and evolutionary studies of *Aquilegia*. *New Phytologist*, *161*(1), 113-120.

Hosouchi, T., Tsuruoka, H., Kotani, H. (2007). Sequencing analysis of chloroplast DNA. Unpublished.

Hu, Z. Y., Hua, W., Huang, S. M., & Wang, H. Z. (2011). Complete chloroplast genome sequence of rapeseed (*Brassica napus L.*) and its evolutionary implications. *Genetic Resources and Crop Evolution*, *58*(6), 875-887.

Hu, S., Sablok, G., Wang, B., Qu, D., Barbaro, E., Viola, R., Li, M., & Varotto, C. (2015). Plastome organization and evolution of chloroplast genes in Cardamine species adapted to contrasting habitats. *BMC Genomics*, *16*(1), 306.

Huang, C. H., Sun, R., Hu, Y., Zeng, L., Zhang, N., Cai, L., Zhang, Q., Koch, M. A., Al-Shehbaz, I., Edger, P. P., Pires, J. C., Tan, D. Y., Zhong, Y., & Ma, H. (2015). Resolution of Brassicaceae phylogeny using nuclear genes uncovers nested radiations and supports convergent morphological evolution. *Molecular Biology and Evolution*, *33*(2), 394-412.

Hunt, B. G., Ometto, L., Wurm, Y., Shoemaker, D., Soojin, V. Y., Keller, L., & Goodisman, M. A. (2011). Relaxed selection is a precursor to the evolution of phenotypic plasticity. *Proceedings of the National Academy of Sciences*, *108*(38), 15936-15941.

Iida, S., Miyagi, A., Aoki, S., Ito, M., Kadono, Y., & Kosuge, K. (2009). Molecular adaptation of *rbcL* in the heterophyllous aquatic plant *Potamogeton*. *PLoS One*, *4*(2), e4633.

Inada, M., Sasaki, T., Yukawa, M., Tsudzuki, T., & Sugiura, M. (2004). A systematic search for RNA editing sites in pea chloroplasts: an editing event causes diversification from the evolutionarily conserved amino acid sequence. *Plant and Cell Physiology*, *45*(11), 1615-1622.

Jansen R. K., Cai Z., Raubeson L. A., Daniell H., dePamphilis C. W.,Leebens-Mack J., Müller K. F., Guisinger-Bellian, M., Haberle, R. C., Hansen, A. K., Chumley, T. M., Lee, S. B., Peery, R., McNeal, J. R., Kuehl, J. V., & Boore, J. L. (2007). Analysis of 81 genes from 64 plastid genomes resolves relationships in angiosperms and identifies genome-scale evolutionary patterns. *Proceedings of the National Academy of Sciences*, *104*(49), 19369-19374.

Jansen, R. K., Raubeson, L. A., Boore, J. L., Chumley, T. W., Haberle, R. C., Wyman, S. K., Alverson, A. J., Peery, R., Herman, S. J., Fourcade, H. M., Kuehl, J. V., McNeal. J.R., Leebens-Mack, J., & Cui L. (2005). Methods for obtaining and analyzing whole chloroplast genome sequences. *Methods in Enzymology*, *395*, 348-384.

Jansen, R. K., Wojciechowski, M. F., Sanniyasi, E., Lee, S. B., & Daniell, H. (2008). Complete plastid genome sequence of the chickpea (*Cicer arietinum*) and the phylogenetic distribution of *rps12* and *clpP* intron losses among legumes (Leguminosae). *Molecular Phylogenetics and Evolution*, *48*(3), 1204-1217.

Janssen, T., & Bremer, K. (2004). The age of major monocot groups inferred from 800+ *rbcL* sequences. *Botanical Journal of the Linnean Society*, *146*(4), 385-398.

Jeffroy, O., Brinkmann, H., Delsuc, F., & Philippe, H. (2006). Phylogenomics: the beginning of incongruence?. *Trends in Genetics*, *22*(4), 225-231.

Johnson, L. A., & Soltis, D. E. (1994). *matK* DNA sequences and phylogenetic reconstruction in Saxifragaceae s. str. *Systematic Botany*, *19*(1),143-156.

Johnston, J. S., Pepper, A. E., Hall, A. E., Chen, Z. J., Hodnett, G., Drabek, J., Lopez, R., & Price, H. J. (2005). Evolution of genome size in Brassicaceae. *Annals of Botany*, *95*(1), 229-235.

Jordon-Thaden, I., Hase, I., Al-Shehbaz, I., & Koch, M. A. (2010). Molecular phylogeny and systematics of the genus *Draba* (Brassicaceae) and identification of its most closely related genera. *Molecular Phylogenetics and Evolution*, *55*(2), 524-540.

Judd, W. S., Sanders, R. W., & Donoghue, M. J. (1994). Angiosperm family pairs: preliminary phylogenetic analyses. *Harvard Papers in Botany*, (5), 1-51.

Kagale, S., Robinson, S. J., Nixon, J., Xiao, R., Huebert, T., Condie, J., Kessler, D., Clarke, W. E., Edger, P.P., Links, M. G., Sharpe, A. G., & Sharpe, A. G. (2014). Polyploid evolution of the Brassicaceae during the Cenozoic era. *The Plant Cell*, *26*(7), 2777-2791.

Källersjö, M., Farris, J. S., Chase, M. W., Bremer, B., Fay, M. F., Humphries, C. J., Petersen, G., Seberg, O., & Bremer, K. (1998). Simultaneous parsimony jackknife analysis of 2538 *rbcL* DNA sequences reveals support for major clades of green plants, land plants, seed plants and flowering plants. *Plant Systematics and Evolution*, *213*(3), 259-287.

Kapralov, M. V., & Filatov, D. A. (2007). Widespread positive selection in the photosynthetic Rubisco enzyme. *BMC Evolutionary Biology*, *7*(1), 1-10.

Katoh, K., Kuma, K. I., Toh, H., & Miyata, T. (2005). MAFFT version 5: improvement in accuracy of multiple sequence alignment. *Nucleic Acids Research*, *33*(2), 511-518.

Kenrick, P., & Crane, P. R. (1997). The origin and early evolution of plants on land. *Nature*, *389*(6646), 33-39.

Khalik, K. A., Van Der Maesen, L. J. G., Koopman, W. J. M., & van den Berg, R. G. (2002). Numerical taxonomic study of some tribes of Brassicaceae from Egypt. *Plant Systematics and Evolution*, *233*(3), 207-221.

Kiemnec, G. L., & McInnis, M. L. (2002). Hoary cress (*Cardaria draba*) root extract reduces germination and root growth of five plant species. *Weed Technology*, *16*(1), 231-234.

Kikuchi, S., Bédard, J., Hirano, M., Hirabayashi, Y., Oishi, M., Imai, M., Takase, M., Ide, T., & Nakai, M. (2013). Uncovering the protein translocon at the chloroplast inner envelope membrane. *Science*, *339*(6119), 571-574.

Kim, K. J., & Jansen, R. K. (1995). *ndhF* sequence evolution and the major clades in the sunflower family. *Proceedings of the National Academy of Sciences*, *92*(22), 10379-10383.

Kim, Y. K., Park, C. W., & Kim, K. J. (2009). Complete chloroplast DNA sequence from a Korean endemic genus, *Megaleranthis saniculifolia*, and its evolutionary implications. *Molecules and Cells*, *27*(3), 365-381.

Kimura, M. (1968). Evolutionary rate at the molecular level. *Nature*, *217*(5129), 624-626.

Kiss, É., Kós, P. B., Chen, M., & Vass, I. (2012). A unique regulation of the expression of the *psbA*, *psbD*, and *psbE* genes, encoding the D1, D2 and cytochrome b559 subunits of the Photosystem II complex in the chlorophyll d containing cyanobacterium *Acaryochloris marina*. *Biochimica et Biophysica Acta -Bioenergetics*, *1817*(7), 1083-1094.

Kluth, C., & Bruelheide, H. (2005). Central and peripheral *Hornungia petraea* populations: patterns and dynamics. *Journal of Ecology*, *93*(3), 584-595.

Knight, R. D., Freeland, S. J., & Landweber, L. F. (2001). A simple model based on mutation and selection explains trends in codon and amino-acid usage and GC composition within and across genomes. *Genome Biology*, *2*(4), 1-13.

Koch, M. A., & Mummenhoff, K. (2006). Editorial: Evolution and phylogeny of the Brassicaceae. *Plant Systematics and Evolution*, *259*(2), 81-83.

Koch, M. A., Dobeš, C., Kiefer, C., Schmickl, R., Klimeš, L., & Lysak, M. A. (2007). Supernetwork identifies multiple events of plastid *trnF* (GAA) pseudogene evolution in the Brassicaceae. *Molecular Biology and Evolution*, *24*(1), 63-73.

Koch, M., Al-Shehbaz, I. A., & Mummenhoff, K. (2003). Molecular systematics, evolution, and population biology in the mustard family (Brassicaceae). *Annals of the Missouri Botanical Garden*, *90*(2), 151-171.

Koch, M., Haubold, B., & Mitchell-Olds, T. (2001). Molecular systematics of the Brassicaceae: evidence from coding plastidic *matK* and nuclear *Chs* sequences. *American Journal of Botany*, *88*(3), 534-544.

Kramer, E. M. (2009). *Aquilegia*: a new model for plant development, ecology, and evolution. *Annual Review of Plant Biology*, *60*, 261-277.

Kramer, E. M., & Hodges, S. A. (2010). *Aquilegia* as a model system for the evolution and ecology of petals. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, *365*(1539), 477-490.

Kramer, E. M., Holappa, L., Gould, B., Jaramillo, M. A., Setnikov, D., & Santiago, P. M. (2007). Elaboration of B gene function to include the identity of novel floral organs in the lower eudicot *Aquilegia*. *The Plant Cell*, *19*(3), 750-766.

Kubo, T., & Newton, K. J. (2008). Angiosperm mitochondrial genomes and mutations. *Mitochondrion*, *8*(1), 5-14.

Kurtz, S., Phillippy, A., Delcher, A. L., Smoot, M., Shumway, M., Antonescu, C., & Salzberg, S. L. (2004). Versatile and open software for comparing large genomes. *Genome Biology*, *5*(2), 816.

Lahti, D. C., Johnson, N. A., Ajie, B. C., Otto, S. P., Hendry, A. P., Blumstein, D. T., Coss, R. G., Donohue, K., & Foster, S. A. (2009). Relaxed selection in the wild. *Trends in Ecology & Evolution*, *24*(9), 487-496.

Lam, V. K., Gomez, M. S., & Graham, S. W. (2015). The highly reduced plastome of mycoheterotrophic *Sciaphila* (Triuridaceae) is colinear with its green relatives and is under strong purifying selection. *Genome Biology and Evolution*, *7*(8), 2220-2236.

Langmead, B., Trapnell, C., Pop, M., & Salzberg, S. L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biology*, *10*(3), 1-10.

Lassmann, T., & Sonnhammer, E. L. (2005). Kalign–an accurate and fast multiple sequence alignment algorithm. *BMC bioinformatics*, *6*(1), 1-9.

Li, L. F., Wang, H. Y., Pang, D., Liu, Y., Liu, B., & Xiao, H. X. (2014). Phenotypic and genetic evidence for ecological speciation of *Aquilegia japonica* and *A. oxysepala*. *New Phytologist*, *204*(4), 1028-1040.

Li, R., Li, Y., Kristiansen, K., & Wang, J. (2008). SOAP: short oligonucleotide alignment program. *Bioinformatics*, *24*(5), 713-714.

Lihová, J., Aguilar, J. F., Marhold, K., & Feliner, G. N. (2004). Origin of the disjunct tetraploid *Cardamine amporitana* (Brassicaceae) assessed with nuclear and chloroplast DNA sequence data. *American Journal of Botany*, *91*(8), 1231-1242.

Lihová, J., Carlsen, T., Brochmann, C., & Marhold, K. (2009). Contrasting phylogeographies inferred for the two alpine sister species *Cardamine resedifolia* and *C. alpina* (Brassicaceae). *Journal of Biogeography*, *36*(1), 104-120.

Lihová, J., Marhold, K., Kudoh, H., & Koch, M. A. (2006). Worldwide phylogeny and biogeography of *Cardamine flexuosa* (Brassicaceae) and its relatives. *American Journal of Botany*, *93*(8), 1206-1221.

Little, M. C., & Hallick, R. B. (1988). Chloroplast *rpoA, rpoB*, and *rpoC* genes specify at least three components of a chloroplast DNA-dependent RNA polymerase active in tRNA and mRNA transcription. *Journal of Biological Chemistry*, *263*(28), 14302-14307.

Liu, Q., & Xue, Q. (2005). Comparative studies on codon usage pattern of chloroplasts and their host nuclear genes in four plant species. *Journal of Genetics*, *84*(1), 55-62.

Magee, A. M., Aspinall, S., Rice, D. W., Cusack, B. P., Sémon, M., Perry, A. S., Stefanović, S., Milbourne, D., Barth, S., Palmer, J. D., Gray, J. C., Kavanagh, T. A., & Wolfe, K. H. (2010). Localized hypermutation and associated gene losses in legume chloroplast genomes. *Genome Research*, *20*(12), 1700-1710.

Maliga, P. (1993). Towards plastid transformation in flowering plants. *Trends in Biotechnology*, *11*(3), 101-107.

Manuell, A., Beligni, M. V., Yamaguchi, K., & Mayfield, S. P. (2004). Regulation of chloroplast translation: interactions of RNA elements, RNA-binding proteins and the plastid ribosome. *Biochemical Society Transactions*, *32*(6), 1135-1135.

Marcus, Y., Altman-Gueta, H., Finkler, A., & Gurevitz, M. (2003). Dual role of cysteine 172 in redox regulation of ribulose 1, 5-bisphosphate carboxylase/oxygenase activity and degradation. *Journal of Bacteriology*, *185*(5), 1509-1517.

Margulis, L. (1970). *Origin of eukaryotic cells: Evidence and research implications for a theory of the origin and evolution of microbial, plant and animal cells on the precambrian Earth*. Yale University Press.

Marhold, K., Lihova, J., Grupe, R., & Neuffer, B. (2002). Natural hybridization in Cardamine (Brassicaceae) in the Pyrenees: evidence from morphological and molecular data. *Botanical Journal of the Linnean Society*, *139*(3), 275-294.

Marini, L., Prosser, F., Klimek, S., & Marrs, R. H. (2008). Water-energy, land-cover and heterogeneity drivers of the distribution of plant species richness in a mountain region of the European Alps. *Journal of Biogeography*, *35*(10), 1826-1839.

Marin-Navarro, J. U. L. I. A., & Moreno, J. (2006). Cysteines 449 and 459 modulate the reduction–oxidation conformational changes of ribulose 1, 5-bisphosphate carboxylase/oxygenase and the translocation of the enzyme to membranes during stress. *Plant, Cell and Environment*, *29*(5), 898-908.

Marshall, D. M., Muhaidat, R., Brown, N. J., Liu, Z., Stanley, S., Griffiths, H., Sage, R. F., & Hibberd, J. M. (2007). *Cleome*, a genus closely related to Arabidopsis, contains species spanning a developmental progression from C3 to C4 photosynthesis. *The Plant Journal*, *51*(5), 886-896.

Martínez-Sánchez, A., Llorach, R., Gil, M. I., & Ferreres, F. (2007). Identification of new flavonoid glycosides and flavonoid profiles to characterize rocket leafy salads (*Eruca vesicaria* and *Diplotaxis tenuifolia*). *Journal of Agricultural and Food Chemistry*, *55*(4), 1356-1363.

Mithen, R., Bennett, R., & Marquez, J. (2010). Glucosinolate biochemical diversity and innovation in the Brassicales. *Phytochemistry*, *71*(17), 2074-2086.

Mitreva, M., Wendl, M. C., Martin, J., Wylie, T., Yin, Y., Larson, A., Parkinson, J., Waterston, R. H., & McCarter, J. P. (2006). Codon usage patterns in Nematoda: analysis based on over 25 million codons in thirty-two species. *Genome Biology*, *7*(8), 1-19.

Mohl, H. V. (1835). Sur la structure et les formes des graines de pollen. *Annales des Sciences Naturelles*. *2*(3), 148-346.

Moore, M. J., Bell, C. D., Soltis, P. S., & Soltis, D. E. (2007). Using plastid genome-scale data to resolve enigmatic relationships among basal angiosperms. *Proceedings of the National Academy of Sciences*, *104*(49), 19363-19368.

Moore, M. J., Dhingra, A., Soltis, P. S., Shaw, R., Farmerie, W. G., Folta, K. M., & Soltis, D. E. (2006). Rapid and accurate pyrosequencing of angiosperm plastid genomes. *BMC Plant Biology*, *6*(1), 1-13.

Moore, M. J., Soltis, P. S., Bell, C. D., Burleigh, J. G., & Soltis, D. E. (2010). Phylogenetic analysis of 83 plastid genes further resolves the early diversification of eudicots. *Proceedings of the National Academy of Sciences*, *107*(10), 4623-4628.

Morton, B. R., & Clegg, M. T. (1993). A chloroplast DNA mutational hotspot and gene conversion in a noncoding region near *rbcL* in the grass family (Poaceae). *Current Genetics*, *24*(4), 357-365.

Müller, A., Diener, S., Schnyder, S., Stutz, K., Sedivy, C., & Dorn, S. (2006). Quantitative pollen requirements of solitary bees: implications for bee conservation and the evolution of bee-flower relationships. *Biological Conservation*, *130*(4), 604-615.

Murrell, B., De Oliveira, T., Seebregts, C., Pond, S. L. K., & Scheffler, K. (2012a). Modeling HIV-1 drug resistance as episodic directional selection. *PLoS Computational Biology*, *8*(5), e1002507.

Murrell, B., Moola, S., Mabona, A., Weighill, T., Sheward, D., Kosakovsky Pond, S. L., & Scheffler, K. (2013). FUBAR: a fast, unconstrained bayesian approximation for inferring selection. *Molecular Biology and Evolution*, *30*(5), 1196-1205.

Murrell, B., Weaver, S., Smith, M. D., Wertheim, J. O., Murrell, S., Aylward, A., Eren, K., Pollner, T., Martin, D. P., Smith, D. M., Scheffler, K., & Kosakovsky Pond, S.L. (2015). Gene-wide identification of episodic selection. *Molecular Biology and Evolution*, *32*(5), 1365-1371.

Murrell, B., Wertheim, J. O., Moola, S., Weighill, T., Scheffler, K., & Pond, S. L. K. (2012b). Detecting individual sites subject to episodic diversifying selection. *PLoS Genetics*, *8*(7), e1002764.

Mutlu, B. (2004). A new species of *Arabis* (Brassicaceae) from inner Anatolia. *Botanical Journal of the Linnean Society*, *145*(2), 251-256.

Nachman, M. W. (2006). Detecting selection at the molecular level. In: *Evolutionary Genetics, Concepts and Case Studies* (eds FoxCW, WolfJB), pp. 103–118. Oxford University Press, Oxford, UK.

Nair, R. R., Nandhini, M. B., Monalisha, E., Murugan, K., Sethuraman, T., Nagarajan, S., Rao, N. S., & Ganesh, D. (2012). Synonymous codon usage in chloroplast genome of *Coffea arabica*. *Bioinformation*, *8*(22), 1096-1104.

Nardi, E. (2011). Diagnoses aquilegiarum novarum in Europa crescentium. *Webbia*, *66*(2), 231-232.

Neckermann, K., Zeltz, P., Igloi, G. L., Kössel, H., & Maier, R. M. (1994). The role of RNA editing in conservation of start codons in chloroplast genomes. *Gene*, *146*(2), 177-182.

Nei, M. (2005). Selectionism and neutralism in molecular evolution. *Molecular Biology and Evolution*, *22*(12), 2318-2342.

Nei, M., & Kumar, S. (2000). *Molecular evolution and phylogenetics*. Oxford university press, New York.

Nickrent, D. L., Parkinson, C. L., Palmer, J. D., & Duff, R. J. (2000). Multigene phylogeny of land plants with special reference to bryophytes and the earliest land plants. *Molecular Biology and Evolution*, *17*(12), 1885-1895.

Nielsen, B. L., Cupp, J. D., & Brammer, J. (2010). Mechanisms for maintenance, replication, and repair of the chloroplast genome in plants. *Journal of Experimental Botany*, *61*(10), 2535-2537.

Odintsova, M. S., & Yurina, N. P. (2005). Genomics and evolution of cellular organelles. *Russian Journal of Genetics*, *41*(9), 957-967.

Ohta, N., Matsuzaki, M., Misumi, O., Miyagishima, S., Nozaki, H., Tanaka, K., Shin-i, T., Kohara, Y. & Kuroiwa, T. (2003). Complete sequence and analysis of the plastid genome of the unicellular red alga *Cyanidioschyzon merolae*. *DNA Research*, *10*(2), 67-77.

Okonechnikov, K., Golosova, O., & Fursov, M. (2012). Unipro UGENE: a unified bioinformatics toolkit. *Bioinformatics*, *28*(8), 1166-1167.

Olejniczak, S. A., Łojewska, E., Kowalczyk, T., & Sakowicz, T. (2016). Chloroplasts: state of research and practical applications of plastome sequencing. *Planta*, *244*(3), 517-527.

Ooka, H., Satoh, K., Doi, K., Nagata, T., Otomo, Y., Murakami, K., Matsubara, K., Osato, N., Kawai, J., Carninci, P., Hayashizaki, Y., Suzuki, K., Kojima, K., Takahara, Y., Yamamoto, K., & Kikuchi, S. (2003). Comprehensive analysis of NAC family genes in *Oryza sativa* and *Arabidopsis thaliana*. *DNA Research*, *10*(6), 239-247.

Ossowski, S., Schneeberger, K., Lucas-Lledó, J. I., Warthmann, N., Clark, R. M., Shaw, R. G., Weigel, D., & Lynch, M. (2010). The rate and molecular spectrum of spontaneous mutations in *Arabidopsis thaliana*. *Science*, *327*(5961), 92-94.

Palmer, J. D., Soltis, D. E., & Chase, M. W. (2004). The plant tree of life: an overview and some points of view. *American Journal of Botany*, *91*(10), 1437-1445.

Parisod, C., & Besnard, G. (2007). Glacial in situ survival in the Western Alps and polytopic autopolyploidy in *Biscutella laevigata* L. (Brassicaceae). *Molecular Ecology*, *16*(13), 2755-2767.

Park, S., Jansen, R. K., & Park, S. (2015). Complete plastome sequence of *Thalictrum coreanum* (Ranunculaceae) and transfer of the *rpl32* gene to the nucleus in the ancestor of the subfamily Thalictroideae. *BMC Plant Biology*, *15*(1), 1-12.

Parks, M., Cronn, R., & Liston, A. (2009). Increasing phylogenetic resolution at low taxonomic levels using massively parallel sequencing of chloroplast genomes. *BMC Biology*, *7*(1), 1-17.

Perný, M., Tribsch, A., Stuessy, T. F., & Marhold, K. (2005). Allopolyploid origin of *Cardamine silana* (Brassicaceae) from Calabria (southern Italy): karyological, morphological and molecular evidence. *Botanical Journal of the Linnean Society*, *148*(1), 101-116.

Petit, R. J., Duminil, J., Fineschi, S., Hampe, A., Salvini, D., & Vendramin, G. G. (2005). Invited review: comparative organization of chloroplast, mitochondrial and nuclear diversity in plant populations. *Molecular Ecology*, *14*(3), 689-701.

Philippe, M., & Ochyra, R. (2004). Occurrence of the moss *Buxbaumia viridis* (Bryopsida, Buxbaumiaceae) in the Tatras National Park (Poland). *Bryological Studies in the Western Carpathians. Sorus, Poznań*, 29-36.

Pond, S. L. K., Frost, S. D., & Muse, S. V. (2005). HyPhy: hypothesis testing using phylogenies. *Bioinformatics*, 21(5), 676-679.

Pond, S. L. K., Posada, D., Gravenor, M. B., Woelk, C. H., & Frost, S. D. (2006). Automated phylogenetic detection of recombination using a genetic algorithm. *Molecular Biology and Evolution*, *23*(10), 1891-1901

Prażmo, W. (1965). Cytogenetic studies on the genus *Aquilegia*, III. Inheritance of the traits distinguishing different complex in the genus *Aquilegia*. *Acta Societatis Botanicorum Poloniae*, *34*(3), 403-437.

Price, D. H., and Gray, M. W. (1998). "Editing of tRNA." *Modification and Editing of RNA*. American Society of Microbiology, 289-305.

Prosser, F., & Bertolli, A. (2007). Nuova combinazione in Brassica (Brassicaceae). *Annali Museo civico Rovereto, Sezione: Archeologia, Storia, Scienze naturali*, *22*, 295-297.

Qiu, Y. L., Li, L., Wang, B., Chen, Z., Knoop, V., Groth-Malonek, M., Dombrovska, O., Lee, J., Kent, L., Rest, J., Estabrook, G. F., Hendrya, T. A., Taylora, D. W., Testab, C. M., Ambrosb, M., Crandall-Stotlerg, B., Duffh, R. J., Stechi, M., Freyi, W., Quandtj, D., & Davisk, C. C. (2006). The deepest divergences in land plants inferred from phylogenomic evidence. *Proceedings of the National Academy of Sciences*, *103*(42), 15511-15516.

Raju, G. P., Rao, P. S., Sekhar, V. C., & Rao, C. S. (2016). In silico study in mitochondrial and chloroplast genomes of plants. *International Journal of Computer Engineering Research*, *6*(2), 8-16.

Rand, D. M., & Kann, L. M. (1996). Excess amino acid polymorphism in mitochondrial DNA: contrasts among genes from *Drosophila*, mice, and humans. *Molecular Biology and Evolution*, *13*(6), 735-748.

Ranwez, V., Harispe, S., Delsuc, F., & Douzery, E. J. (2011). MACSE: Multiple Alignment of Coding SEquences accounting for frameshifts and stop codons. *PLoS One*, *6*(9), e22594.

Raubeson, L. A., Peery, R., Chumley, T. W., Dziubek, C., Fourcade, H. M., Boore, J. L., & Jansen, R. K. (2007). Comparative chloroplast genomics: analyses including new sequences from the angiosperms *Nuphar advena* and *Ranunculus macranthus*. *BMC Genomics*, *8*(1), 1-27.

Regvar, M., Eichert, D., Kaulich, B., Gianoncelli, A., Pongrac, P., & Vogel-Mikuš, K. (2013). Biochemical characterization of cell types within leaves of metal-hyperaccumulating *Noccaea praecox* (Brassicaceae). *Plant and Soil*, *373*(1), 157-171.

Regvar, M., Vogel-Mikuš, K., Kugonič, N., Turk, B., & Batič, F. (2006). Vegetational and mycorrhizal successions at a metal polluted site: Indications for the direction of phytostabilisation?. *Environmental Pollution*, *144*(3), 976-984.

Reith, M., & Munholland, J. (1995). Complete nucleotide sequence of the *Porphyra purpurea* chloroplast genome. *Plant Molecular Biology Reporter*, *13*(4), 333-335.

Roberts, K. (ed.) (2007) *Handbook of Plant Science*. Chichester, UK: John Wiley & Sons.

Rochaix, J. D. (2005). The three genomes of Chlamydomonas. *Photosynthesis Research*, *73*(1), 285-293.

Rodríguez-Moreno, L., González, V. M., Benjak, A., Martí, M. C., Puigdomènech, P., Aranda, M. A., & Garcia-Mas, J. (2011). Determination of the melon chloroplast and mitochondrial genome sequences reveals that the largest reported mitochondrial genome in plants contains a significant amount of DNA having a nuclear origin. *BMC Genomics*, *1*(12), 1-14.

Rogers, M. B., Gilson, P. R., Su, V., McFadden, G. I., & Keeling, P. J. (2007). The complete chloroplast genome of the chlorarachniophyte *Bigelowiella natans*: evidence for independent origins of chlorarachniophyte and euglenid secondary endosymbionts. *Molecular Biology and Evolution*, *24*(1), 54-62.

Rohde, W., Gramstat, A., Schmitz, J., Tacke, E., & Prüfer, D. (1994). Plant viruses as model systems for the study of non-canonical translation mechanisms in higher plants. *Journal of General Virology*, *75*(9), 2141-2149.

Ronquist, F., & Huelsenbeck, J. P. (2003). MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics*, *19*(12), 1572-1574.

Roquet, C., Coissac, É., Cruaud, C., Boleda, M., Boyer, F., Alberti, A., Gielly, L., Taberlet, P., Thuiller, W., Van Es, J., & Lavergne, S. (2016). Understanding the evolution of holoparasitic plants: the complete plastid genome of the holoparasite *Cytinus hypocistis* (Cytinaceae). *Annals of Botany*, *118*(5), 885-896.

Rothfels, C. J., Larsson, A., Kuo, L. Y., Korall, P., Chiou, W. L., & Pryer, K. M. (2012). Overcoming deep roots, fast rates, and short internodes to resolve the ancient rapid radiation of eupolypod II ferns. *Systematic Biology*, *61*(3), 490-509.

Rousseau-Gueutin, M., Huang, X., Higginson, E., Ayliffe, M., Day, A., & Timmis, J. N. (2013). Potential functional replacement of the plastidic acetyl-CoA carboxylase subunit (*accD*) gene by recent transfers to the nucleus in some angiosperm lineages. *Plant Physiology*, *161*(4), 1918-1929.

Sablok, G., Nayak, K. C., Vazquez, F., & Tatarinova, T. V. (2011). Synonymous codon usage, GC3, and evolutionary patterns across plastomes of three pooid model species: emerging grass genome models for monocots. *Molecular Biotechnology*, *49*(2), 116-128.

Saito, N., Tatsuzawa, F., Toki, K., Shinoda, K., Shigihara, A., & Honda, T. (2011). The blue anthocyanin pigments from the blue flowers of *Heliophila coronopifolia L*. (Brassicaceae). *Phytochemistry*, *72*(17), 2219-2229.

Sasaki, T., Yukawa, Y., Miyamoto, T., Obokata, J., & Sugiura, M. (2003). Identification of RNA editing sites in chloroplast transcripts from the maternal and paternal progenitors of tobacco (*Nicotiana tabacum*): comparative analysis shows the involvement of distinct trans-factors for *ndhB* editing. *Molecular Biology and Evolution*, *20*(7), 1028-1035.

Sato, S., Nakamura, Y., Kaneko, T., Asamizu, E., & Tabata, S. (1999). Complete structure of the chloroplast genome of *Arabidopsis thaliana*. *DNA Research*, *6*(5), 283-290.

Savolainen, V., Chase, M. W., Hoot, S. B., Morton, C. M., Soltis, D. E., Bayer, C., Fay, M. F., De Bruijn, A. Y., Sullivan, S., & Qiu, Y. L. (2000). Phylogenetics of flowering plants based on combined analysis of plastid *atpB* and *rbcL* gene sequences. *Systematic Biology*, *49*(2), 306-362.

Scheen, A. C., Elven, R., & Brochmann, C. (2002). A molecular morphological approach solves taxonomic controversy in arctic *Draba* (Brassicaceae). *Canadian Journal of Botany*, *80*(1), 59-71.

Schranz, M. E., & Mitchell-Olds, T. (2006). Independent ancient polyploidy events in the sister families Brassicaceae and Cleomaceae. *The Plant Cell*, *18*(5), 1152-1165.

Schwienbacher, E., Navarro-Cano, J. A., Neuner, G., & Erschbamer, B. (2011). Seed dormancy in alpine species. *Flora-Morphology, Distribution, Functional Ecology of Plants*, *206*(10), 845-856.

Sebastia, M. T., Casals, P., Vojniković, S., Bogunić, F., & Beus, V. (2005). Plant diversity and soil properties in pristine and managed stands from Bosnian mixed forests. *Forestry*, *78*(3), 297-303.

Sen, L., Fares, M. A., Liang, B., Gao, L., Wang, B., Wang, T., & Su, Y. J. (2011). Molecular evolution of *rbcL* in three gymnosperm families: identifying adaptive and coevolutionary patterns. *Biology Direct*, *6*(1), 29-29.

Sen, L., Fares, M. A., Su, Y. J., & Wang, T. (2012). Molecular evolution of *psbA* gene in ferns: unraveling selective pressure and co-evolutionary pattern. *BMC Evolutionary Biology*, *12*(1), 1-13.

Shaw, J., Shafer, H. L., Leonard, O. R., Kovach, M. J., Schorr, M., & Morris, A. B. (2014). Chloroplast DNA sequence utility for the lowest phylogenetic and phylogeographic inferences in angiosperms: The tortoise and the hare IV. *American Journal of Botany*, *101*(11), 1987-2004.

Shendure, J., & Ji, H. (2008). Next-generation DNA sequencing. *Nature Biotechnology*, *26*(10), 1135-1145.

Shikanai, T. (2015). RNA editing in plants: Machinery and flexibility of site recognition. *Biochimica et Biophysica Acta-Bioenergetics*, *1847*(9), 779-785.

Shimda, H., & Sugiuro, M. (1991). Fine structural features of the chloroplast genome: comparison of the sequenced chloroplast genomes. *Nucleic Acids Research*, *19*(5), 983-995.

Shinozaki, K., Ohme, M., Tanaka, M., Wakasugi, T., Hayashida, N., Matsubayashi, T., Zaita, N., Chunwongse, J., Obokata, J., Yamaguchi-Shinozaki, K., Ohto., Torazawa, K., Meng B. Y., Sugita, M., Deno, H., Kamogashira, T., Yamada, K., Kusuda, J., Takaiwa, F., Kato, A., Tohdoh, N., Shimada, H., & Sugiura, M. (1986). The complete nucleotide sequence of the tobacco chloroplast genome. *Plant Molecular Biology Reporter*, *4*(3), 111-148.

Small, R. L., Ryburn, J. A., Cronn, R. C., Seelanan, T., & Wendel, J. F. (1998). The tortoise and the hare: choosing between noncoding plastome and nuclear *Adh* sequences for phylogeny reconstruction in a recently diverged plant group. *American Journal of Botany*, *85*(9), 1301-1315.

Šmarda, P., Bureš, P., Horová, L., Foggi, B., & Rossi, G. (2008). Genome size and GC content evolution of *Festuca*: ancestral expansion and subsequent reduction. *Annals of Botany*, *101*(3), 421-433.

Smith, D. R., Lee, R. W., Cushman, J. C., Magnuson, J. K., Tran, D., & Polle, J. E. (2010). The *Dunaliella salina* organelle genomes: large sequences, inflated with intronic and intergenic DNA. *BMC Plant Biology*, *10*(1), 1-14.

Smith, S. A., & Donoghue, M. J. (2008). Rates of molecular evolution are linked to life history in flowering plants. *Science*, *322*(5898), 86-89.

Smith, M. D., Wertheim, J. O., Weaver, S., Murrell, B., Scheffler, K., & Kosakovsky Pond, S. L. (2015). Less is more: an adaptive branch-site random effects model for efficient detection of episodic diversifying selection. *Molecular Biology and Evolution*, *32*(5), 1342-1353.

Soltis, D. E., Soltis, P. S., & Tate, J. A. (2004). Advances in the study of polyploidy since plant speciation. *New Phytologist*, *161*(1), 173-191.

Soltis, D. E., Soltis, P. S., Chase, M. W., Mort, M. E., Albach, D. C., Zanis, M., Savolainen, V., Hahn, W. H., Hoot, S.B., Fay, M. F., Axtell, M., Swensen, S. M., Prince, L. M., Kress, W. J., Nixon, K. C., & Farris, J. S. (2000). Angiosperm phylogeny inferred from 18S rDNA, *rbcL*, and *atpB* sequences. *Botanical Journal of the Linnean Society*, *133*(4), 381-461.

Soltis, P. S., Soltis, D. E., Savolainen, V., Crane, P. R., & Barraclough, T. G. (2002). Rate heterogeneity among lineages of tracheophytes: integration of molecular and fossil data and evidence for molecular living fossils. *Proceedings of the National Academy of Sciences*, *99*(7), 4430-4435.

Soltis, P. S., Soltis, D. E., Wolf, P. G., Nickrent, D. L., Chaw, S. M., & Chapman, R. L. (1999). The phylogeny of land plants inferred from 18S rDNA sequences: pushing the limits of rDNA signal?. *Molecular Biology and Evolution*, *16*(12), 1774-1784.

Stadermann, K. B., Weisshaar, B., & Holtgräwe, D. (2015). SMRT sequencing only *de novo* assembly of the sugar beet (*Beta vulgaris*) chloroplast genome. *BMC Bioinformatics*, *16*(1), 1-10.

Stamatakis, A., Hoover, P., & Rougemont, J. (2008). A rapid bootstrap algorithm for the RAxML web servers. *Systematic Biology*, *57*(5), 758-771.

Stern, A., Doron-Faigenboim, A., Erez, E., Martz, E., Bacharach, E., & Pupko, T. (2007). Selecton 2007: advanced models for detecting positive and purifying selection using a Bayesian inference approach. *Nucleic Acids Research*, *1*, 6.

Stevens, P.F. (2001 onwards). Angiosperm Phylogeny Website. Version 12, July 2012 [and more or less continuously updated since]. http://www.mobot.org/MOBOT/research/APweb/

Stöver, B. C., & Müller, K. F. (2010). TreeGraph 2: combining and visualizing evidence from different phylogenetic analyses. *BMC Bioinformatics*, *11*(1), 7.

Strasburger, E. (1884). Neuere untersuchungen über den befruchtungsvorgang bei den phanerogamen als grundlage für eine theorie der zeugung.Jena, Gustav Fischer.

Straub, S. C., Parks, M., Weitemier, K., Fishbein, M., Cronn, R. C., & Liston, A. (2012). Navigating the tip of the genomic iceberg: Next-generation sequencing for plant systematics. *American Journal of Botany*, *99*(2), 349-364.

Su, Y. J., & Wang, T. (2008). Adaptive evolution in the rubisco large subunits of epiphytic ferns (Polypodiaceae): identification of positively selected sites. *Acta Scientiarum Naturalium Universitatis Sunyatseni*, *5*, 016.

Sudhani, H. P. (2012). Control of the activity of Rubisco from *Chlamydomonas reinhardtii* through the redox state of its cysteine residues. (Doctoral dissertation, Universitat de València).

Sueoka, N. (1964). On the evolution of informational macromolecules. *Evolving Genes and Proteins. Academic Press: New York*, 479-96.

Surina, B. (2002). Phytogeographical differentiation in the Dinaric fir-beech forest (*Omphalodo-Fagetum s. lat.*) of the western part of the Illyrian floral province. *Acta Botanica Croatica*, *61*(2), 145-178.

Swanson, W. J., Nielsen, R., & Yang, Q. (2003). Pervasive adaptive evolution in mammalian fertilization proteins. *Molecular Biology and Evolution*, *20*(1), 18-20.

Sweeney, P. W., & Price, R. A. (2000). Polyphyly of the genus *Dentaria* (Brassicaceae): evidence from *trnL* intron and *ndhF* sequence data. *Systematic Botany*, *25*(3), 468-478.

Swofford, D. L. (2003). PAUP*: phylogenetic analysis using parsimony, version 4.0 b10. Sinauer Associates, Sunderland, Massachusetts.

Szmidt, A. E., Aldén, T., & Hällgren, J. E. (1987). Paternal inheritance of chloroplast DNA in *Larix*. *Plant Molecular Biology*, *9*(1), 59-64.

Szpara, M. L., Tafuri, Y. R., Parsons, L., Shamim, S. R., Verstrepen, K. J., Legendre, M., & Enquist, L. W. (2011). A wide extent of inter-strain diversity in virulent and vaccine strains of alphaherpesviruses. *PLoS Pathogens*, *7*(10), e1002282.

Takagi, S. (2003). Actin-based photo-orientation movement of chloroplasts in plant cells. *Journal of Experimental Biology*, *206*(12), 1963-1969.

Takenaka, M., Verbitskiy, D., Zehrmann, A., Härtel, B., Bayer-Császár, E., Glass, F., & Brennicke, A. (2014). RNA editing in plant mitochondria-connecting RNA target sequences and acting proteins. *Mitochondrion*, *19*, 191-197.

Takenaka, M., Zehrmann, A., Verbitskiy, D., Härtel, B., & Brennicke, A. (2013). RNA editing in plants and its evolution. *Annual Review of Genetics*, *47*, 335-352.

Talavera, G., & Castresana, J. (2007). Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. *Systematic Biology*, *56*(4), 564-577.

Tamura, M. (1993). Ranunculaceae. *Flowering Plants· Dicotyledons*. Springer Berlin Heidelberg. 563-583

Tamura, M. N., Yamashita, J., Fuse, S., & Haraguchi, M. (2004). Molecular phylogeny of monocotyledons inferred from combined analysis of plastid *matK* and *rbcL* gene sequences. *Journal of Plant Research*, *117*(2), 109-120.

Tang, J., Xia, H. A., Cao, M., Zhang, X., Zeng, W., Hu, S., Tong, W., Wang, J., Wang, J., Yu, J., Yang, H., & Zhu, L. (2004). A comparison of rice chloroplast genomes. *Plant Physiology*, *135*(1), 412-420.

Tangphatsornruang, S., Uthaipaisanwong, P., Sangsrakru, D., Chanprasert, J., Yoocha, T., Jomchai, N., & Tragoonrung, S. (2011). Characterization of the complete chloroplast genome of *Hevea brasiliensis* reveals genome rearrangement, RNA editing sites and phylogenetic relationships. *Gene*, *475*(2), 104-112.

Tatarinova, T. V., Alexandrov, N. N., Bouck, J. B., & Feldmann, K. A. (2010). GC 3 biology in corn, rice, sorghum and other grasses. *BMC Genomics*, *11*(1), 308.

Tremetsberger, K., König, C., Samuel, R., Pinsker, W., & Stuessy, T. F. (2002). Infraspecific genetic variation in *Biscutella laevigata* (Brassicaceae): new focus on Irene Manton's hypothesis. *Plant Systematics and Evolution*, *233*(3), 163-181.

Turgay, O. C., Görmez, A., & Bilen, S. (2012). Isolation and characterization of metal resistant-tolerant rhizosphere bacteria from the serpentine soils in Turkey. *Environmental Monitoring and Assessment*, *184*(1), 515-526.

Urbanska, K. M. (1997). Restoration ecology research above the timberline: colonization of safety islands on a machine-graded alpine ski run. *Biodiversity and Conservation*, *6*(12), 1655-1670.

Voelckel, C., Borevitz, J. O., Kramer, E. M., & Hodges, S. A. (2010). Within and between whorls: comparative transcriptional profiling of Aquilegia and Arabidopsis. *PLoS One*, *5*(3), e9735.

Wan, X. F., Xu, D., Kleinhofs, A., & Zhou, J. (2004). Quantitative relationship between synonymous codon usage bias and GC composition across unicellular genomes. *BMC Evolutionary Biology*, *4*(1), 1-11.

Wang, M., Kapralov, M. V., & Anisimova, M. (2011). Coevolution of amino acid residues in the key photosynthetic enzyme Rubisco. *BMC Evolutionary Biology*, *11*(1), 1-12.

Warwick, S. I., Francis, A., & Al-Shehbaz, I. A. (2006). Brassicaceae: species checklist and database on CD-Rom. *Plant Systematics and Evolution*, *259*(2-4), 249-258.

Warwick, S. I., Sauder, C. A., Al-Shehbaz, I. A., & Jacquemoud, F. (2007). Phylogenetic relationships in the tribes Anchonieae, Chorisporeae, Euclidieae, and Hesperideae (Brassicaceae) based on nuclear ribosomal ITS DNA sequences. *Annals of the Missouri Botanical Garden*, *94*(1), 56-78.

Wertheim, J. O., Murrell, B., Smith, M. D., Pond, S. L. K., & Scheffler, K. (2015). RELAX: detecting relaxed selection in a phylogenetic framework. *Molecular Biology and Evolution*, *32*(3), 820-832.

Werz, G., & Kellner, G. (1968). Molecular characteristics of chloroplast DNA of *Acetabularia* cells. *Journal of Ultrastructure Research*, *24*(1), 109-115.

Whatley, J. M., & Whatley, F. R. (1981). Chloroplast evolution. *New Phytologist*, *87*(2), 233-247.

Whittall, J. B., & Hodges, S. A. (2007). Pollinator shifts drive increasingly long nectar spurs in columbine flowers. *Nature*, *447*(7145), 706-709.

Whittall, J. B., Voelckel, C., Kliebenstein, D. J., & Hodges, S. A. (2006). Convergence, constraint and the role of gene expression during adaptive radiation: floral anthocyanins in *Aquilegia*. *Molecular Ecology*, *15*(14), 4645-4657.

Wicke, S., Müller, K. F., Quandt, D., Bellot, S., & Schneeweiss, G. M. (2016). Mechanistic model of evolutionary rate variation en route to a nonphotosynthetic lifestyle in plants. *Proceedings of the National Academy of Sciences of the United States of America*, *113*(32), 9045-9050.

Wicke, S., Schäferhoff, B., dePamphilis, C. W., & Müller, K. F. (2014). Disproportional plastome-wide increase of substitution rates and relaxed purifying selection in genes of carnivorous Lentibulariaceae. *Molecular Biology and Evolution*, *31*(3), 529-545.

Wink, M., & Mohamed, G. I. (2003). Evolution of chemical defense traits in the *Leguminosae*: mapping of distribution patterns of secondary metabolites on a molecular phylogeny inferred from nucleotide sequences of the *rbcL* gene. *Biochemical Systematics and Ecology*, *31*(8), 897-917.

Wu, F. H., Kan, D. P., Lee, S. B., Daniell, H., Lee, Y. W., Lin, C. C., Lin, N. S., & Lin, C. S. (2009). Complete nucleotide sequence of *Dendrocalamus latiflorus* and *Bambusa oldhamii* chloroplast genomes. *Tree Physiology*, *29*(6), 847-856.

Wu, Z., Gui, S., Quan, Z., Pan, L., Wang, S., Ke, W., Liang, D., & Ding, Y. (2014). A precise chloroplast genome of *Nelumbo nucifera* (Nelumbonaceae) evaluated with Sanger, Illumina MiSeq, and PacBio RS II sequencing platforms: insight into the plastid evolution of basal eudicots. *BMC Plant Biology*, *14*(1), 289-289.

Wyman, S. K., Jansen, R. K., & Boore, J. L. (2004). Automatic annotation of organellar genomes with DOGMA. *Bioinformatics*, *20*(17), 3252-3255.

Xu, C., Cai, X., Chen, Q., Zhou, H., Cai, Y., & Ben, A. (2011). Factors affecting synonymous codon usage bias in chloroplast genome of *Oncidium Gower Ramsey*. *Evolutionary Bioinformatics*, 7, 271-278.

Yang, Y., Zhou, T., Duan, D., Yang, J., Feng, L., & Zhao, G. (2016). Comparative analysis of the complete chloroplast genomes of five *Quercus* species. *Frontiers in Plant Science*, *7*, 959.

Yang, Z., Nielsen, R., Goldman, N., & Pedersen, A. M. K. (2000). Codon-substitution models for heterogeneous selection pressure at amino acid sites. *Genetics*, *155*(1), 431-449.

Zerbino, D. R., & Birney, E. (2008). Velvet: algorithms for *de novo* short read assembly using de Bruijn graphs. *Genome Research*, *18*(5), 821-829.

Zhang, J., Nielsen, R., & Yang, Z. (2005). Evaluation of an improved branch-site likelihood method for detecting positive selection at the molecular level. *Molecular Biology and Evolution*, *22*(12), 2472-2479.

Zhong, B., Xi, Z., Goremykin, V. V., Fong, R., Mclenachan, P. A., Novis, P. M., Davis, C. C., & Penny, D. (2014). Streptophyte algae and the origin of land plants revisited using heterogeneous models with three new algal chloroplast genomes. *Molecular Biology and Evolution*, *31*(1), 177-183.

Zhong, B., Yonezawa, T., Zhong, Y., & Hasegawa, M. (2009). Episodic evolution and adaptation of chloroplast genomes in ancestral grasses. *PLoS One*, *4*(4), e5297.

Zhou, M., Long, W., & Li, X. (2008). Patterns of synonymous codon usage bias in chloroplast genomes of seed plants. *Forestry Studies in China*, *10*(4), 235-242.

Zunk, K., Mummenhoff, K., & Hurka, H. (1999). Phylogenetic relationships in tribe Lepidieae (Brassicaceae) based on chloroplast DNA restriction site variation. *Canadian Journal of Botany*, *77*(10), 1504-1512.

Acknowledgements

Firstly, I would like to express my sincere gratitude to my advisor Dr. Claudio Varotto for his unwavering support, collegiality, and mentorship over the entire period of my PhD study and related research; Besides my advisor, I would like to thank Prof. Dario Leister for the support from the PhD program in Biology of Munich University (Ludwig-Maximilians Universität), which provides me great chances to communicate and learn.

My sincere thanks also go to Dr. Li Mingai and Enrico Barbaro, who provided me an opportunity to join their team as intern, and who gave access to the laboratory and research facilities. Without their precious support it would not have been possible to conduct this research.

I thank my fellow labmates (Dr. José Manuel Carli, Dr. Wang Bo, Dr. Hu Shiliang, Dr. Mastaneh Ahrar, Dr. Wei Yu, Dr. Michele Poli, Dr. Fu Yuan and Jike Wuhe) for the stimulating discussions, and for all the fun I have had in the last four years.

I wish to express my sincere thanks to— Fondazione Edmund Mach di San Michele all'Adige for providing me with necessary facilities; I am deeply grateful for the financial support from the Chinese scholarship council (CSC), which provided me the opportunity to study aborad; I would like to extend my thanks to those who paved the path of phylogenetics and evolution study in green plants, also to experts and referees for their precise and accurate comments for prompting me to prepare a better thesis.

Last but not the least, I would like to thank my family: my parents, my brother and his wife for supporting me spiritually throughout writing this thesis and my life in general. Also thanks to my cute niece, whose smile give me power when I feel tired. A very special gratitude goes out to my boyfriend who have supported me along the way, even if I am too busy and have no time to get along with him.

Thanks for all your encouragement!

Curriculum vitae

Name: Huan Li
Gender: Female

QUALIFICATIONS

Familiar with Common Bioinformatics Tools
Computer Skill and Programming: Proficient in Perl and Linux system commands, also Microsoft Office Word, Excel and PowerPoint
Adept in communication and cooperation, have strong team spirit.

PROFESSIONAL EXPERIENCE

Dealing with the next generation sequencing data of chloroplast DNA: Sequenced more than 1,100 representative species of wild species from an alpine flora. Used different approaches to assemble the reads, align contigs through best-fit references, annotate plastomes and construct phylogenetic relationships. Evolutionary studies of natural selection were applied on plenty of protein coding sequences across families at gene-wide, site and branch levels. I further assessed whether these results might be affected by RNA-editing at different levels. Additionally, several selected families with a sufficient number of representatives and variation in altitudinal distribution were tested whether omega rates significantly differed between low and high elevation taxa by relaxed selection. This study not only provides a paradigm on how to proceed towards the full elucidation of the evolutionary relationships among various biological species in the tree of life, but also provides a deeper understanding of the evolutionary patterns of the plastome across families which co-evolved in the same floristic assemblage. (10/2013--present)

Predicting the protein-protein interactions network of pathogens: Used interolog, domain combination and K-Nearest Neighbors verification methods to predict the protein-protein interactions and established the network by Cytoscape program. Detailed analyses in the flagellar synthesis and the chemotaxis system, the signal transduction system in *Xoo* PXO99^{A} provided valuable clues to explore the pathogenicity and metabolic regulation of this plant pathogen. (09/2010-06/2013)

"Chip, computer, crop" International Academic Symposium, Huazhong Agricultural University: Took part in the curriculum about analysis of the gene chip data with respect to crop research, and as the group representative to analyze results and share learning experience through a speech. (09/2011)

The 9th International Bioinformatics Workshop (IBW), the Fourth Military

Medical University: Took part in the curriculum of "Biological Sequence Analysis and Systems Biology Probabilistic Graphical Model" for five days and the seminar of "Bioinformatics International Symposium" for three days. (07/2011)

Research Assistant, Plant Tissue Laboratory of Jianghan University: Applied and completed the academic science and technology of student project "Established the regeneration system of virus-free seedlings of *Lily*". Detoxification technology was used to obtain tissue culture materials from the illness of leaf, then filtered out the virus-free plantlets through virus detection, finally tuber-propagated and planted them. (06/2008-06/2009)

EDUCATION

Bachelor of Science (2006-2010)
Major: Biotechnology
College of Life Sciences, Jianghan University, Wuhan, P.R. China

Master of Science (2010-2013)
Major: Microbiology
College of Life Science and Technology, Huazhong Agricultural University, Wuhan, P.R. China

Doctor of Philosophy (2013--present)
Major: Bioinformatics
Faculty of Biology, Ludwig Maximilian University of Munich, Munich, Germany
Department of Biodiversity and Molecular Ecology, Research and Innovation Centre, Fondazione Edmund Mach, Trento, Italy

AWARDS

2013-2017 Scholarship of China Scholarship Council
2011-2012 Excellent Merit Postgraduate of Huazhong Agricultural University
2010-2011 Excellent Postgraduate Cadre of Huazhong Agricultural University
2008-2009 Second scholarship of Jianghan University
2007-2008 First scholarship of Jianghan University
Excellent Student Leader of Jianghan University
2006-2007 Second scholarship of Jianghan University
Excellent Student Leader of Jianghan University

PUBLICATION

Jing Guo[#], Huan Li[#], Ji-Wei Chang, Yang Lei, Sen Li, Ling-Ling Chen (2013). Prediction and characterization of protein-protein interaction network in *Xanthomonas oryzae* pv. *oryzae* PXO99[A]. *Research in Microbiology*, *164*(10), 1035-1044.