University of Reading

Towards a monograph in *Narcissus*; problems and challenges in the *N. minor* complex

A thesis submitted by

Jordan Bilsborrow

For the Degree of Doctor of Philosophy

University of Reading Herbarium,

School of Biological Sciences, University of Reading

January 2019

Declaration

I confirm that this is my own work and the use of all material from other sources has been properly and fully acknowledged.

Jordan Bilsborrow

Narcissus cyclamineus DC.

Acknowledgments

Firstly, I would like to thank the Royal Horticultural Society and Celia Baxter for providing the scholarship to make this project possible. I am grateful to my supervisors A. Culham and J. David for all the advice, support and encouragement that they have provided.

I am deeply grateful to my (almost) supervisor K. Könyves for his valuable support, training and assistance during the project and M. Christodoulou for statistics counselling, and continued support.

Special thanks are required to all of the collaborators and contacts in Morocco, Portugal, and Spain who have assisted in helping to organise fieldwork and collection permits, of which there are too many to mention. I am extremely grateful to all friends and colleagues that helped with collections by assisting field excursions: S. Jury, I. Maruri, M. Pagola, T. Pitman, F. Amigo-Vázquez, and S. Tejero-García. I would like to thank Brian Duncan for welcoming me to his daffodil collection, and providing a vast amount of knowledge about the study group. The Culham Research group have been helpful as a source of support, interesting discussion and amusement.

I am indebted to both my parents, brother, Paula and Frank for their love and support, and more importantly than all for listening even when they had no clue what I was saying. Last but by no means least… ขอบคุณศวิตา ที่คอยเป็นกำลังใจให้ผมตลอดมา จนมาถึงวันนี้ได้

Abstract

Narcissus is a genus of well-known and widely grown garden plants that shows conflict between morphological and molecular taxa. This is compounded by complex breeding programmes that have given rise to over 30,000 registered cultivars, themselves difficult to tell apart. Current identification methods require meticulous study of morphological characters that are available only seasonally. Many of the modern cultivars are a result of hybridization of species from subsection *Pseudonarcissi* including the *Narcissus minor* group. Population-level sampling was used to study genetic variation within the *N. minor* group and to explore the congruence of genetic and taxonomic-morphological boundaries. More than 600 individuals were collected from 56 populations across the geographic range of the species, and a reference living collection for *Narcissus* established. A combination of microsatellite markers, Sanger and Next-Generation sequencing were used to generate data. Strong geographic structure to population genetics of the *N. minor* group is revealed using microsatellite analysis and plastid DNA sequencing, that is incongruent with the current taxonomic treatment. Further to this, microsatellite markers were applied to the study of cultivar identification. Successful transferability of microsatellites to cultivated daffodils ranged from 39-100%. Despite the complex interbreeding history of daffodil cultivars genetic patterns recovered were able to distinguish the cultivars studied. As expected, many of these cultivars did not form genetic groups that were congruent with the horticultural daffodil Divisions in which those cultivars occur. However, the utility of microsatellites is demonstrated, distinguishing most cultivars and highlighting mislabelled stocks. Ten new complete plastomes for *Narcissus* are presented, as a first step in quantifying levels of differences across the genus and the variation among these genomes is compared. The data gathered from genome skimming will provide a valuable resource for marker development for future classification of species and cultivar identification. Overall, both microsatellite and genomic approaches have been shown to resolve taxa however the taxa recovered do not match those currently recognised using morphology.

Table of Contents

Chapter 1　General Introduction

1.1 Taxonomy and species concepts

Daffodils are an economically important group of plants that have many cultural links (Dweck, 2002). Despite the economic and cultural significance, the species boundaries within *Narcissus* are often poorly characterised, with the last comprehensive taxonomic revision was published in 1875 (Baker, 1875). Many species have been described in the time since, and only regional treatments have been published (Webb, 1980, Aedo, 2013, Fennane, Ibn Tattou, et al., 2014). This has given rise to a range in the number of accepted species from 16 (Baker, 1875) to 160 (Haworth, 1831). The range of species described is partly a product of no definitive definition of a species. Garnett & Christidis (2017) suggested standardising species lists to promote taxonomic stability for the purpose of species conservation. This would reduce the large range of species names being accepted in genera such as *Narcissus* however, the idea is fundamentally flawed. Taxonomy, like any other branch of science, uses new methods on new data to test hypotheses, and as a result re-evaluates the taxonomic concepts. Limiting taxonomic studies could actually hamper conservation, as it is not possible to conserve what you do not know. Hey et al. (2003) outlined that species related research and conservation should proceed even with the ambiguity of species concepts.

For daffodils, the morphological species concept has been the most widely applied, using morphology to characterise species for the use in keys (Fernandes, 1968) and floras (Webb, 1980, Aedo, 2010). Criticism of the morphological species concept is usually centred on inaccurate character definition, which could introduce ambiguity to the work (Christodoulou and Culham, 2018). The prevalent hybridisation between *Narcissus* species can obscure the morphological boundaries (Marques, Feliner, et al., 2010), and the wide range in ploidy adds further complexity to classifying species (Fernandes, 1975). In the past 20 years phylogenetic study of *Narcissus* has resulted in patterns of variation that conflict with previously defined species limits that have been based upon morphological data (Graham and Barrett, 2004, Santos-Gally, Vargas, et al., 2012, Marques, Aguilar, et al., 2017, Könyves, David, et al., 2019).

One of the issues with *Narcissus* species classification has included recognising variation at the taxonomic level of either species, sub-species, or variety. An example of this is *N. asturiensis,* which has been recognised at species and subspecies level, *N. minor* subsp. *asturiensis.* This has depended upon the breadth of the species concept used to describe morphological

variation. The issue of recognising species should therefore be overcome using a combination of datasets to elucidate species boundaries, as suggested by Doyen and Slobodchikoff (1974).

The aim of this thesis is to evaluate the variation within the *Narcissus minor* group, and closely related taxa in subsection *Pseudonarcissi* in order to elucidate species boundaries. Subsection *Pseudonarcissi* has made a large contribution to horticultural daffodils, therefore understanding the wild variation patterns from which cultivars were originally selected has implications for the identification of daffodil cultivars. This thesis also aims to demonstrate if a molecular means can be established for daffodil cultivar identification and to use Next-Generation sequencing to find regions of the genome which can be used for classification and as a genetic resource for marker development.

1.2 Taxonomy of *Narcissus*

The genus *Narcissus* L. is placed within the family Amaryllidaceae J. St.-Hil. (Saint-Hilaire, 1805), a family which consists largely of bulbous geophytes, with basal linear leaves. Amaryllidaceae floral parts are in multiples of 3 (3+3 tepals, 3+3 stamens, 3 carpellate ovary), with tepals fused at the base, forming a tube subtended by a spathe. The family contains many horticulturally popular genera beside *Narcissus* such as: *Allium*, *Amaryllis*, *Hippeastrum*, *Galanthus*, *Leucojum*, *Nerine* and *Sternbergia* (Heywood, Brummit, et al., 2007). Amaryllidaceae is divided into three subfamilies based upon molecular evidence (A.P.G III, 2009), of which *Narcissus* belongs to Amaryllidoideae Burnett, this subfamily is further divided into 14 tribes (Chase, Reveal, et al., 2009). *Narcissus* is placed within *Narcisseae* along with *Sternbergia*.

Narcissus was formally named by Linnaeus in *Species Plantarum* in 1753, with the genus type *N. poeticus* chosen in 1913 (Britton and Brown, 1913). Initially many of the sections presently included in *Narcissus* were separate genera: *Ajax*, *Corbularia*, *Hermione*, *Queltia, Ganymedes* (Haworth, 1831). The genera were later recognised at lower taxonomic ranks within *Narcissus* (Baker, 1875). Since the cytological work of Abilio Fernandes there has been general consensus that the genus comprises two subgenera: *Narcissus* (x=7), and *Hermione* (x=5) (Fernandes, 1951). This was later disputed in the account of Flora Europaea since chromosome counts were the only character to uniquely separate the subgenera at that time (Webb, 1978), and subsequent cytological work of DNA contents identified overlap between the two subgenera (Zonneveld, 2008). However, more recent phylogenetic analysis of the genus resolved both

subgenera as monophyletic groups based upon chloroplast (*matK*, *ndhF*) and nuclear (ITS) sequence data (Graham and Barrett, 2004, Marques, Aguilar, et al., 2017, Könyves, David, et al., 2019), with the exception of two closely related taxa; *N. tortifolius* and *N. dubius* which are ancient allopolyploids (Zonneveld, 2008).

The genus is further divided into sections (Table 1-1) with both morphological (Fernandes, 1968, Webb, 1978, Blanchard, 1990, Mathew, 2002, Aedo, 2013) and cytological (Fernandes, 1951, 1975) studies providing evidence of 10 sections (Table 1-1). The ten sections are recognisable based upon morphological and phenological characters (Blanchard, 1990). Subsection ranks have also been utilised to recognise the variation below the rank of section based upon morphological (Webb, 1980) and phylogenetic study (Marques, Aguilar, et al., 2017). Phylogenetic studies encompassing chloroplast, mitochondrial and nuclear gene data found that only two sections *Apodanthi* and *Serotini* are monophyletic, excluding the monotypic section *Aurelia* (Graham and Barrett, 2004, Santos-Gally, Vargas, et al., 2012, Marques, Aguilar, et al., 2017). Therefore, several sections were split or changed taxonomic rank, as shown in Table 1-1, this has led to 14 sections being proposed (Marques, Aguilar, et al., 2017).

Table 1-1 Details of *Narcissus* sections with number of species compiled from Kington (2008), and compared to Marques et al. (2017).

Section	No. of species	No. of subspecies	No. of varieties	Section in Marques et al. (2017)	Ploidy of section
Subgenus *Hermione*					
Aurelia	1	0	0	*Aurelia*	×2, ×4
Serotini	2	0	2	*Serotini*	×2, ×6
Tazettae	16	5	8	*Tazettae, Dubii, Angustifolii*	×2, ×3, ×4, ×6
Subgenus *Narcissus*					
Apodanthi	6	2	1	*Apodanthi*	×2
Braxireon	1	0	0	*Braxireon*	×2, ×4
Bulbocodii	11	8	14	*Bulbocodii, Meridionalis*	×2-8
Ganymedes	1	2	5	*Ganymedes*	×2
Jonquillae	8	1	8	*Jonquillae, Juncifolii*	×2-4
Narcissus	2	8	0	Subsect. *Narcissus*	×2, ×3
Pseudonarcissi	37	6	25	Subsect. *Pseudonarcissi, Nevadensis*	×2, ×3, ×4, ×6

Sections such as *Apodanthi*, *Ganymedes*, and *Tazettae* have provided a model system to explore the evolution of stylar polymorphism and breeding systems, using phylogenetic and molecular data (Arroyo and Barrett, 2000, Graham and Barrett, 2004, Pérez-Barrales, Vargas, et al., 2004, Hodgins and Barrett, 2007, Santos-Gally, Gonzalez-Voyer, et al., 2013, Simon-Porcar, Pico, et al., 2015). This has led to a deeper understanding of the variation within these sections, floral diversification, and the variation in pollinators. Within *Narcissus* three different style-length polymorphisms occur, at least 12 species exhibit stigma height dimorphism, as well as the two forms of heterostyly; distyly (*N. albimarginatus*), and tristyly (*N. triandrus*) (Thompson, 2005, Barranco, Arroyo, et al., 2019). Many species in *Narcissus* have been recorded to be pollinated by generalist insects such as bees, hawkmoths, diurnal moths, and pierid butterflies (Pérez, Vargas, et al., 2003). The floral variation in widespread daffodil species such as *N. papyraceus* has been correlated with shifts in pollinator; populations predominantly pollinated by syrphid flies have a shorter tube and smaller tepals, compared with populations pollinated by moths (Pérez-Barrales, Arroyo, et al., 2007). It has been proposed that pollinators provide the selective pressure for floral convergence in *Narcissus*, that has led to morphological similarities in floral traits (Arroyo and Barrett, 2000).

Compounding the issue of using morphology to classify daffodil species is the prevalent hybridisation and subsequent introgression that is exhibited in daffodils (Marques, Nieto Feliner, et al., 2012), evident by the polyphyly of some sections (Marques, Aguilar, et al., 2017). The molecular patterns of the phylogenetic tree do not appear to corroborate the morphological based separation. For example, there are discrepancies between floral morphology of hoop-petticoat daffodils and phylogeny (Figure 1-1), which could be due to hybridisation of the two lineages (Könyves, David, et al., 2019).

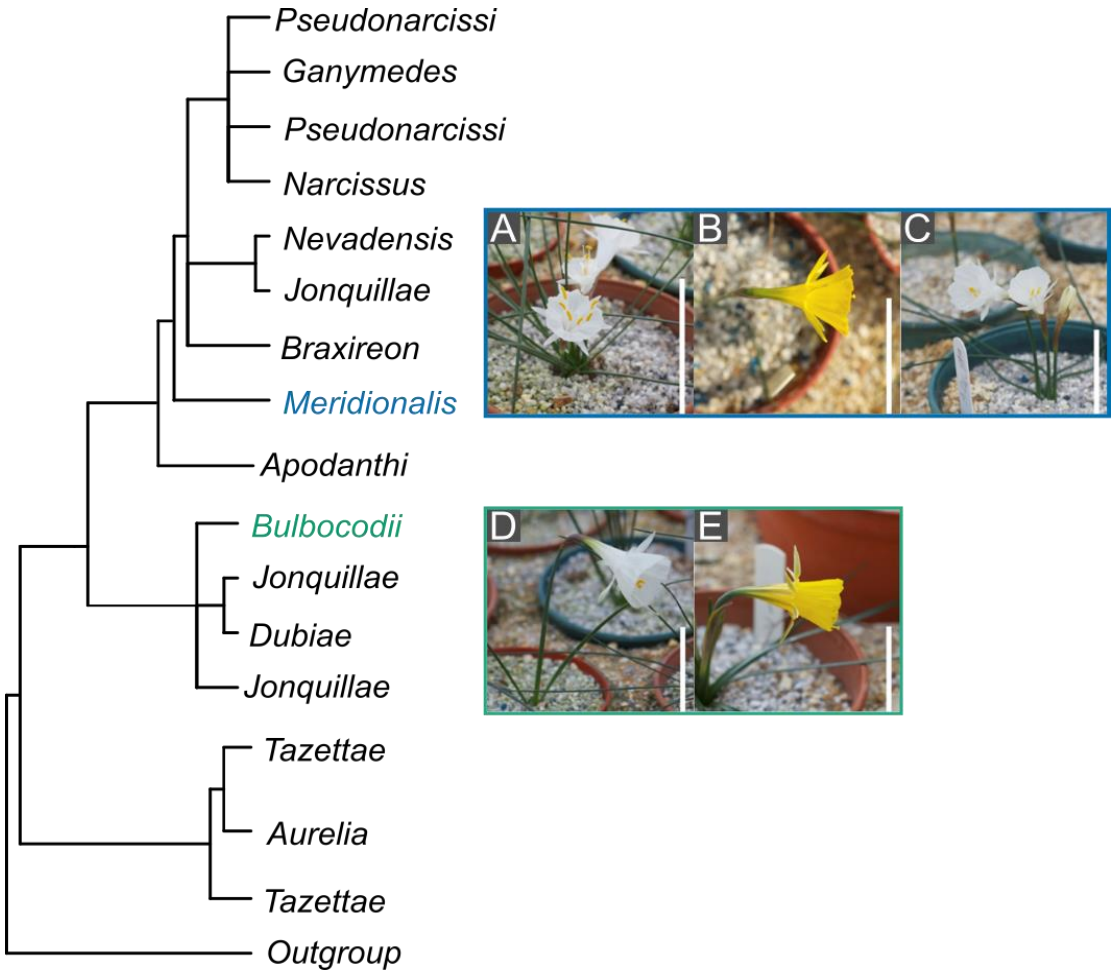


Figure 1-1 Conflict between floral morphology and cpDNA phylogeny in hoop-petticoat daffodils. Images in blue correspond to section *Meridionalis*, images in green correspond to section *Bulbocodii*. A=*N. romieuxii*; B=*N.obesus*; C=*N. cantabricus*; D=*N. cantabricus*; E=*N. bulbocodium.*

Hybridisation within the genus has also led to offspring recruitment of different pollinators, and subsequent isolation from their progenitor species. Shifts in pollinators have been recorded in two hybrids; *N.* × *perezlarae*, and *N.* × *alentejanus*, both pollinated by ants, rather

than bees or butterflies (Marques, Jürgens, et al., 2016). The prevalent hybridisation led to hybrid sections being suggested, shown in Table 1-2, although these have not been widely accepted to date. Increasing the complexity of classification is the contribution polyploidy has made to speciation across the genus (Fernandes, 1975), with taxa ranging from diploid to nonoploid (Zonneveld, 2008). The wide range of morphological variation within taxa, in particular with varying ploidy (e.g. *N. bicolor* and *N. abscissus*), could have contributed to taxonomic inflation (Jiménez, Sánchez-Gómez, et al., 2009).

Table 1-2 Proposed hybrid sections for *Narcissus* (Fernández Casas, 1983, 1984a, 2005, 2013, LLoret i Sabaté and Fernández Casas, 2009, Fernández Casas and Uribe-Echebarría Díaz, 2011).

***Narcissus* hybrid section**	**Combination**	**Protologue**
× *Anni* Fernández Casas	*Angustifolii* × *Serotini*	Fontqueria 56(32):298 (2013)
× *Apocissi* Fernández Casas & Uribe-Echevarría Díaz	*Apodanthi* × *Pseudonarcissi*	Fontqueria 56(18):159-160 (2011)
× *Apomedes* Fernández Casas & Susanna	*Apodanthi* × *Ganymedes*	Anales Jard.Bot.Madrid 38(1):39 (1981)
× *Aurettae* Fernández Casas	*Aurelia* × *Tazettae*	Fontqueria 56(32):298 (2013)
× *Braster* Fernández Casas	*Braxireon* × *Chloraster*	Fontqueria 56(32):299 (2013)
× *Braxireotini* Fernández Casas	*Braxireon* × *Serotini*	Fontqueria 56(32):299 (2013)
× *Bulbisii* Fernández Casas	*Bulbocodii* × *Pseudonarcissi*	Fontqueria 6:36 (1984)
× *Bulbomedes* Fernández Casas	*Bulboodii* × *Ganymedes*	Fontqueria 55(34):270 (2005)
× *Bulbominei* Fernández Casas	*Bulbocodii* × *Cyclaminei*	Fontqueria 56(32):299 (2013)
× *Bulboquillae* Fernández Casas	*Bulbocodii* × *Jonquillae*	Fontqueria 6:37 (1984)
× *Chlorustifolii* Fernández Casas	*Chloraster* × *Seroangustifolii*	Fontqueria 56(32):299 (2013)
× *Dubiquillae* Fernández Casas	*Dubii* × *Jonquillae*	Fontqueria 6:39 (1984)
× *Ganycissi* Fernández Casas	*Ganymedes* × *Pseudonarcissi*	Fontqueria 6:39 (1984)
× *Ganyquillae* Fernández Casas	*Ganymedes* × *Jonquillae*	Fontqueria 56(32):300 (2013)
× *Jonicissus* Fernández Casas	*Jonquillae* × *Narcissus*	Fontqueria 55(34):270 (2005)
× *Jonissi* Fernández Casas	*Jonquillae* × *Pseudonarcissi*	Fontqueria 6:37-39 (1984)
× *Jonquizettae* Fernández Casas	*Jonquillae* × *Tazettae*	Fontqueria 6:39 (1984)
× *Narcisettae* Fernández Casas	*Narcissus* × *Tazettae*	Fontqueria 6:39 (1984)
× *Nassi* Fernández Casas	*Narcissus* × *Pseudonarcissi*	Fontqueria 6:39 (1984)
× *Pseudosettae* Fernández Casas	*Pseudonarcissi* × *Tazettae*	Fontqueria 6:39 (1984)
× *Dubizetae* Fernández Casas	*Dubii* × *Tazettae*	Fontqueria 56(1) (2009)
× *Apocodii* Fernández Casas	*Apodanthi* × *Bulbocodii*	Fontqueria 3:23 (1983)
× *Ponsii-Sorollae* Fernández Casas	*Jonquillae* × *Ganymedes*	Fontqueria 3:27 (1983)
× *Pujolii* Fernández Casas	*Jonquillae* × *Dubii*	Fontqueria 3:29 (1983)

Further contributing to the complexity in classifying *Narcissus* is the close links the plants have to cultivation (Webb, 1978), many taxa have been described from cultivated plants or plants with unknown provenance (Mathew, 2002). Daffodils have been in cultivation for centuries (Lobel, 1570, van de Passe, 1614), and many classifications of wild species stem from a horticultural interest (Pugsley, 1933, Blanchard, 1990). Furthermore, many protologues have also provided insufficient data with which to separate taxa, e.g. *N. jacetanus* Fern. Casas (Fernández Casas, 1984b), and have to date not been critically assessed in order to gain wider acceptance.

1.3 Subsection *Pseudonarcissi*

Subsection *Pseudonarcissi* includes the trumpet daffodils, characterised by a corona which is usually as long as the tepals. The trumpet daffodils are amongst the most speciose in the genus and are distributed throughout Europe (Figure 1-2), the centre for diversity is the Iberian Peninsula. There are no subsection *Pseudonarcissi* taxa distributed in northern Africa. Plants within this subsection occur at a range of altitudes from near sea level to up to 2,000 m (Blanchard, 1990).

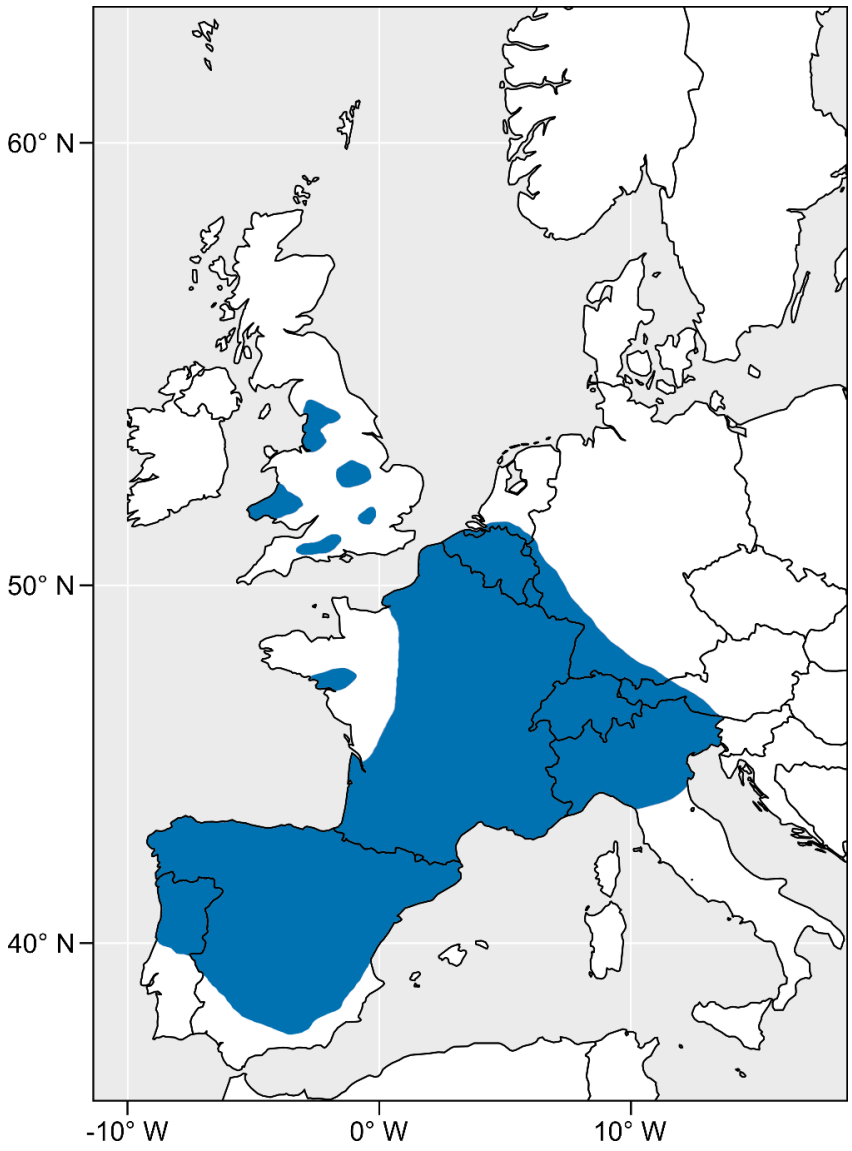


Figure 1-2 Distribution of *Narcissus* section *Pseudonarcissi*. Redrawn from Blanchard (1990).

Subsection *Pseudonarcissi* was originally published in Redouté, Les Liliacées 8 (De Candolle, 1816). The first binomial name for the subsection, *Narcissus pseudonarcissus* L. was described in *Species Plantarum* (Linnaeus 1753). However, varieties of trumpet daffodils date back to before the 16th century, detailed in 1570 (Lobel, 1570). Even before this time the plants had been known throughout western Europe (Coates, 1956). The close links *N. pseudonarcissus* shares with horticulture is highlighted by the Linnaean type specimen that has a double tepal and/or corona and is therefore likely to be from cultivated origin (Figure 1-3).

Figure 1-3 Digitised image of the type specimen of *Narcissus pseudo-narcissus* L. (LINN 412.2) in the Linnean Society of London Herbarium (**LINN**).

Trumpet daffodils have made a large contribution to cultivated daffodils, being involved in the origins of many cultivars in several of the horticultural Divisions (Rivera Nuñez, Obón De Castro, et al., 2003). Their use in cultivation adds complexity to classification of the wild species, as many species descriptions and classifications stem from horticultural interests, and the wild provenance of species is often unknown (Blanchard, 1990).

The section has been described as a separate genus, *Ajax* (Haworth, 1831), used in previous taxonomic treatments including Pugsley's (1933) monograph of the trumpet daffodils. However, the genus *Ajax* was treated as a section of *Narcissus* by Baker (1875), which has been generally accepted since. Section *Pseudonarcissi* initially included all the trumpet daffodils based upon morphological characters as the corona is usually as long as the tepals (Blanchard, 1990). However, cytological evidence separates the trumpet daffodils distributed in southern

Spain into section *Nevadensis* (Zonneveld, 2008), with phylogenetic support (Marques, Aguilar, et al., 2017), and therefore section *Pseudonarcissi* now includes all trumpet daffodils with a solitary flower. Other features of the section include anthers which are always shorter than the style are uniseriate and attached to the filaments at one end rather than centrally (Blanchard, 1990).

A wide range of taxa have been described in section *Pseudonarcissi*, depending on the species concept used to classify them, from 13 (Webb, 1978) to 37 (Mathew, 2002) taxa. Since Pugsley's (1933) monograph of *Ajax* many species descriptions have been published within section *Pseudonarcissi* (Rivera Nuñez, Obón De Castro, et al., 2003), although thorough taxonomic investigation has not taken place for the species to gain wider acceptance. Section *Pseudonarcissi* is not monophyletic, with species within the section found within three clades (Graham and Barrett, 2004, Santos-Gally, Vargas, et al., 2012, Könyves, David, et al., 2019). As *N. poeticus* (section *Narcissus*) is nested within section *Pseudonarcissi* the change in taxonomic rank to subsection *Pseudonarcissi* within section *Narcissus* was suggested by Marques, Aguilar, et al. (2017).

The early cytological analysis found *Narcissus* subsection *Pseudonarcissi* to have a base chromosome number of n=7 (Fernandes, 1951). The majority of taxa studied within the section are diploid. The only known exceptions according to Zonneveld (2008) are *N. bicolor* (2n=6x=42), *N. nobilis* (2n=4x=28) and *N. pseudonarcissus* which has both diploid and hexaploid plants (2n=2x=14 and 2n=6x=42). The varying ploidy contributes to the difficulty in classifying taxa of subsection *Pseudonarcissi.* Species within the *N. minor* group are diploid apart from a tetraploid form of *N. asturiensis*, subspecies *villarvildensis* (Zonneveld, 2008).

1.4 *Narcissus minor* group

The *N. minor* group is comprised of several small stature taxa of subsection *Pseudonarcissi.* The species within the group correspond to one of the three *Pseudonarcissi* phylogenetic clades described above, along with *N. cyclamineus* DC. Pugsley's monograph of *Ajax* (1933) split the genus into section *Cyclaminopsis* and section *Pseudo-Narcissus*, further classifying taxa of section *Pseudo-Narcissus* into series. Six species were classified into series *Minores*; *N. asturiensis* (Jord.) Pugsley, *N. lagoi* Merino, *N. minor* L., *N. pumilus* Salisb., *N. nanus* (Haw.)

Spach, and *N. parviflorus* (Jord.) Pugsley. Taxa from the study group included in this thesis are shown in Figure 1-4.

Figure 1-4 Photographs representing the taxa examined in this thesis, and the populations they were collected from (see Appendix A) One unit of the scale bar equals one centimetre.

Series have not been used in subsequent classifications, although the series *Minores* is still recognised as the *Narcissus minor* group, a morphologically distinct set of taxa within subsection *Pseudonarcissi*. Taxa of the *N. minor* group have been in cultivation for centuries with *N. minor* illustrations appearing in Jardin du Roi Henry IV along with illustrations of

N. cyclamineus (Vallet, 1608), and complexities have arisen through the provenance of material for species descriptions (Pugsley, 1933).

Section *Cyclaminopsis* as recognised by Pugsley (1933) includes *N. cyclamineus* which has been classified within subsection *Pseudonarcissi*, it is morphologically distinct and can be easily separated from other subsection *Pseudonarcissi* taxa due to the reflexed tepals not found in other trumpet daffodils (Webb, 1978). The species was classified in subsection *Reflexi* in Flora Europaea, based upon this morphological character alone (Webb, 1980). In numerous morphological investigations into *Narcissus*, *N. cyclamineus* has been recognised (Fernandes, 1968, Blanchard, 1990, Mathew, 2002, Aedo, 2013). Cytological study has also identified the species from closely related taxa (Fernandes, 1975, Zonneveld, 2008). However, phylogenetic study based upon plastid gene sequences were unable to distinguish the species from taxa within the *N. minor* group (Graham and Barrett, 2004, Könyves, David, et al., 2019).

Narcissus cyclamineus is IUCN red listed, and threatened from habitat destruction through timber plantation and illegal plant collections (Navarro, Ayensa, et al., 2012). *Narcissus cyclamineus* was considered extinct in the wild until rediscovery in 1885 (Blanchard, 1990). The species has a limited distributed in north-west Spain and northern Portugal, although its historic distribution is difficult to ascertain.

The remaining species within the *N. minor* group are more challenging to separate. The origins of *N. minor* were originally unknown, although the species was found in the wild in the early 20th century (Pugsley, 1933) and has been widely included in morphological and cytological classification since (Fernandes, 1968, 1975, Blanchard, 1990, Aedo, 2013). The distinction between *N. minor* and *N. asturiensis* is not clear, *N. minor* is a larger plant than *N. asturiensis*; however, the variation within *N. asturiensis* means there is an overlap in continuous morphological data (Blanchard, 1990). It has been proposed that *N. minor* is of hybrid origin between *N. asturiensis* and *N. pseudonarcissus* (Mathew, 2002). Both species are distributed across the north of the Iberian Peninsula, from the Spanish Pyrenees south-west towards central Portugal. The difficulty separating the two species is reflected in the Flora Iberica account, as *N. asturiensis* is classified as a subspecies of *N. minor* (Aedo, 2013). The illegal collection and habitat destruction of *N. asturiensis* through agricultural use means the species is also IUCN red listed, although the record is data deficient (Caldas and Moreno Saiz, 2011).

More recently, *N. jacetanus* has been described from north east Spain (Fernández Casas, 1984b) distributed from the Spanish Pyrenees to the Basque Country. The original description does not fully characterise the distinguishing characters, only providing continuous morphological data, such as a 2-5 mm long pedicel, and a 28-32 mm long flower (Fernández Casas, 1984). This increases the complexity of separating *N. jacetanus* from previously described species within the *N. minor* group, unless using a geographic split. Populations of *N. jacetanus* are often reported to only grow on limestone (Zonneveld, 2008), although the species has been found occupying similar habitats to *N. asturiensis* (Fernández Casas, 1990). Since the publication of *N. jacetanus* morphological classifications have accepted the species (Blanchard, 1990, Mathew, 2002). However, the Flora Iberica account classifies *N. jacetanus* as a synonym of *N. minor* subsp. *minor* (Aedo, 2013). Phylogenetic analysis has also failed to separate *N. jacetanus* from both *N. asturiensis* and *N. minor* (Graham and Barrett, 2004, Marques, Aguilar, et al., 2017).

The delimitation of *N. lagoi* is also complex, the species has been described from one occurrence near Lugo, Galicia (Merino, 1909) and has not been rediscovered since, this could be due to poor species characterisation. Despite this, the acceptance of the species is mixed based upon morphological classification, the species is accepted in the Flora Europaea account of *Narcissus* (Webb, 1980) and by Mathew (2002), classified as a variety of *N. asturiensis* (Fernandes, 1951), and synonymised as *N. minor* subsp. *asturiensis* (Aedo, 2013). Due to the poor record and characterisation of the species, it has not been sampled in any phylogenetic study to date.

1.5 Horticultural taxonomy & cultivar identification

1.5.1 Cultivar Definition and Naming

"A cultivar as a taxon is an assemblage of plants that (a) has been selected for a particular character or a combination of characters, and (b) remains distinct, uniform, and stable in these characters when propagated by appropriate means." (Brickell, Alexander, et al., 2016)

While this may seem a simple and straightforward definition, the groups of plants named as cultivars can vary substantially in their means of propagation and their inherent underlying genetic variation, as outlined in Table 1-3.

Table 1-3 Types of cultivars based on Culham and Grant (1999) and Brickell, Alexander et al. (2016).

Vegetatively propagated	Genetically uniform	**Clones (Art. 2.6)**	selections arisen from breeding programmes or from wild populations, or chance seedlings propagated vegetatively	e.g. *Gerbera* 'Delphi'
		Topophysic clones or Sports (Art. 2.7)	mutants of somaclonal variation	e.g. *Abies koreana* 'Prostrate Beauty'
		Cyclophysic clones (Art. 2.8)	cultivars derived from a particular phase of a plant's growth cycle	e.g. *Hedera helix* 'Arborescens'
		Viral or other microbial infections (Art. 2.9 & 2.10)	combinations of plants with microbes or viruses, which result in changes in appearance or vigour e.g. witches' broom	e.g. *Picea abies* 'Little Gem'
		Mutation-chimaeras (Art. 2.10)	plants whose characteristics are affected by a proportion of mutant tissue	e.g. *Pelargonium* 'Mrs. Pollock'
		Graft-chimaeras (Art. 2.11)	graft-hybrids that combine the tissue of multiple plants	e.g. + *Laburnocytisus* 'Adamii'
Sexually propagated	Genetically mixed individuals	**Chance seedlings (Art. 2.12)**	seeds derived from uncontrolled pollination	e.g. *Viola* 'Penny Black'
		Topovariant (Art. 2.13)	seeds collected from a particular provenance on more than one occasion	e.g. *Picea abies* of Dutch provenance
		Plants of a line (Art. 2.14)	cultivars arising by repeated self-fertilization or inbreeding	e.g. *Helianthus annuus* 'HA306'
		Plants of a multiline (Art. 2.15)	cultivars originating from repeated inbreeding involving multiple related lines	e.g. T*rifolium repens* 'Star'
		F_1 hybrids (Art. 2.16)	first generation hybrids of two inbred lines	e.g. *Lycopersicon esculentum* 'Cristal'

Table 1-4 (cont.) Types of cultivars based on Culham and Grant (1999) and Brickell, Alexander et al. (2016).

Vegetatively or sexually propogated	Undefined	**Plants arising through ploidy level change (Art. 2.17)**	chromosome doubling in the germ line (the same genes but twice as many)	e.g. *Lilium* 'Tetra Brandywine'
		Agamospermous plants (Art. 2.18)	cultivars propagated from seeds without fertilization	e.g. *Poa pratensis* 'Baron'
		Gemovars	cultivars developed through genetic engineering	e.g. *Petunia* 'African Sunset'

This fundamental variation in the underlying genetic structure of different kinds of cultivars raises further challenges to the consistent identification of them. Cultivars propagated clonally will be highly genetically uniform among individuals and therefore are ideally suited to definition using DNA based marker systems. However, seed raised cultivars are not perfectly uniform in their genetics and any DNA based system of identification must include a sufficient understanding and knowledge of the variation that will occur. F_1 hybrids will show a combination of DNA markers derived from the inbred lines used for F_1 seed production.

The genetic variation between cultivars is essentially genetic variation between individuals. Studying cultivars is therefore most similar to population genetics, in its scope and sampling. Cultivars inherently harbour some of the genetic variation that was present in natural populations, as they have arisen from wild plants, however far back in time this may be. To explain the genetic variation found between and within cultivars accurately, it is essential to understand the extent of variation in the wild and use this as a reference. For example, identifying cultivars belonging to genera where evolutionary processes, or the lack of them, lead to diverse morphological but limited genetic variation, or genetic variation that does not conform to species boundaries is immensely challenging. In biology, such problems are seen in DNA barcoding studies where sometimes a barcode gap is absent (e.g. *Salix*; Percy, Argus, et al., 2014). On the other hand, lack of knowledge of the natural genetic variation in genera where morphological variation is limited, but have well defined genetic structure (e.g. *Cyclamen*; Könyves and Culham, 2014) can lead to underestimating the genetic variation that can be present in cultivars.

Horticultural taxonomy involves the naming and classification of cultivated plants, regardless of the type of cultivar. The International Code of Nomenclature for Cultivated Plants governs the use of the cultivar name and delimitation (Brickell, Alexander, et al., 2016). However, the genus and species from which the cultivar originates is governed by the International Code of Nomenclature for algae, fungi, and plants (Turland, Wiersema, et al., 2018). This adds further complexity to horticultural taxonomy especially considering the ICN outdates the ICNCP, and therefore cultivated plant names were governed by the ICN before 1952, or were given ungoverned by a code.

A reliable form of identification is required for cultivar names to be consistent for repeatable identification. Accurate cultivar identification is particularly vital in vast supply chains of crops were correctly labelled stocks are a legal requirement and an essential part of the ornamental trade, as the cultivar name links the characters and appearance the grower expects. Consumers drive new types of cultivars with ornamental value, such as flower characteristics or colour variation, and therefore the life span of ornamental cultivars is often much shorter than in crops (Yagi, 2018). This has led to vast numbers of cultivars in certain groups of plants with ornamental value. The registration of new cultivars is becoming increasing difficult because new plant selections have traits that can either be hard to describe or harder to prove as distinct from others.

1.5.2 Cultivar Identification Methods

Cultivar identification traditionally relies upon meticulous morphological study, comparing specimens to photographs, herbarium voucher material (at herbaria specifically holding cultivar material, such as **WSY**) or plants in living collections (such as Brogdale Collections for fruit cultivars). This often requires experts in that particular group of cultivars, relying upon years of practical experience. The issue of identification is compounded in groups of plants with few consistent landmarks to examine, and with similar morphologies. This has been highlighted in recent studies using fruit such as apples or peaches for cultivar identification (Christodoulou, Battey, et al., 2018, Yang, Zhang, et al., 2019). Linear and geometric morphometric techniques have been used to classify 27 apple cultivars with an accuracy of 72.6% and 66.7% respectively (Christodoulou, Battey, et al., 2018). Furthermore, machine learning approaches using shape successfully classified 90.7% of apricot specimens however; this was limited to only four apricot

cultivars (Yang, Zhang, et al., 2019). Morphological characters have been applied to identify cultivars, usually limited to small cultivar groups, such as 6 self-sterile bramble cultivars (Kostamo, Toljamo, et al., 2013). However, similar applications to groups such as dahlias, with more than 50,000 cultivars (McClaren, 2009) is impossible. As the number of cultivars increases within a plant group the distinguishing morphological parameters are often lacking, since the cultivars are very closely related.

Using morphology for identification in groups with large numbers of cultivars also heavily relies upon seasonality, such as flower or fruit characters. It is unfeasible to identify daffodil cultivars from leaf and bulb characters alone, and therefore identification is limited to when the plant is in flower. However, a molecular means for cultivar identification is not reliant on flowering time, as DNA can be acquired at any stage of growth.

1.5.3 DNA Fingerprinting

The application of molecular tools for cultivar identification provide an approach to quickly characterise and identify plants before they have reached flowering maturity. For population genetics, DNA fingerprinting methods are most suitable, including: RAPD, AFLP, ISSRs, SSRs, discussed in detail by Culham & Grant (1999). The use of most of these techniques has not declined over in the 20 years since (

Figure 1-5), as their suitability to research questions involving populations or closely related organisms (e.g. cultivars) does not diminish. The use of some fingerprinting methods has increased with the availability of automated fragment analysis (SSRs/microsatellites), and greater access to larger libraries of genetic data (SNPs, SSRs/microsatellites).

The most suitable markers for cultivars are the most rapidly evolving ones, typically deriving from microsatellites or SNPs, the latter becoming more widely available due to the uptake of next-generation sequencing methods, that can identify large numbers of individual markers (Davey, Hohenlohe, et al., 2011).

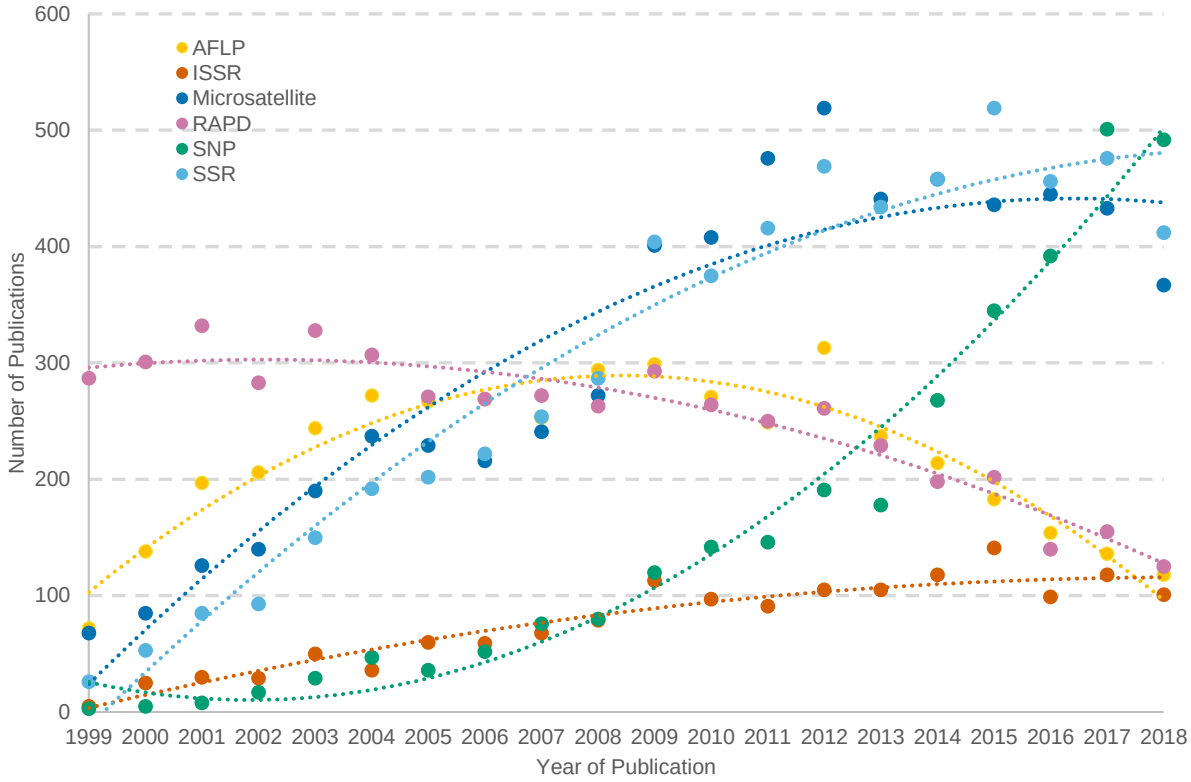


Figure 1-5 Trends in number of plant sciences publications for different DNA fingerprinting methods. SSRs and microsatellites are different terms for the same DNA fingerprinting method, but they are not used consistently in publications. Data collected from Web of Science on 03/09/2019.

Of the fingerprinting methods, SSRs (also known as microsatellites) are more frequently used in population level studies of plants than they were 20 years ago and their application to cultivars has now become both practical and cost-effective once initial libraries are set up (Zalapa, Cuevas, et al., 2012). However, microsatellites are not without limitations. There can be limited transferability of microsatellites between closely related species (Sinama, Dubut, et al., 2011), this could be problematic in plant genera with a large number of artificial crosses amongst species. The issue of cross-amplification is being overcome by the ability to develop much larger libraries. Next-generation sequencing allows microsatellites to be developed across multiple samples or species consequently reducing the cost and time of development, while increasing the transferability by selecting microsatellites that are present in all or most of the studied species/taxa/cultivars (Chapman, 2015, White, Doo, et al., 2016).

A technique hardly known in 20 years ago, SNPs have increased in popularity in the last five years in particular, and has become the most widely used DNA fingerprinting method. However, their application is currently limited to few ornamental plant genera (Ye, Cai, et al., 2016, Chao, Chen, et al., 2018, Lu, Liu, et al., 2018, Chong, Su, et al., 2019, Makki, Saeedi, et al., 2019).

1.5.4 DNA Barcoding

DNA barcoding can be a powerful tool to identify species (Moritz and Cicero, 2004). Standard barcode sequences allow large databases to be constructed, against which any new sample can subsequently be matched. The Barcode for Life Consortium has lead the way in this endeavour and has recommended two regions of chloroplast DNA as the markers for plants: *matK* and *rbcL* (CBOL Plant Working Group, 2009). These regions were chosen due to the successful cross amplification of the *rbcL* region in most plants sampled and the less reliable amplification but higher discriminating power of *matK*. These regions work at different levels, *rbcL* is most suited to identify families and genera, while *matK*, being more variable, ideally providing identification at species level. Empirical tests suggest plant DNA barcoding however has limitations (Li, Feng, et al., 2014, Guo, Huang, et al., 2016). Amplifying and sequencing the *matK* region can be challenging, with successful PCR and sequencing of ~70% of Angiosperms sampled although further primers improving this by approximately 10% (Hollingsworth, Graham, et al., 2011) have now been developed. DNA barcoding can lead to the identification of species groups rather than a single species, where DNA variation amongst species is limited (Stallman, Funk, et al., 2019). Moreover, both official barcode regions are found in the chloroplast genome, which is uniparentally inherited and therefore unequivocal identification of hybrids is impossible.

One way to overcome this is to use nuclear regions alongside *matK* and *rbcL*. A nuclear region proposed is the Internal Transcribed Spacer (ITS) region of the nuclear ribosomal repeat (Hollingsworth, Graham, et al., 2011), which can be successful in resolving species level taxonomy (Baldwin, Sanderson, et al., 1995). ITS however is not without limitations (Feliner and Rosselló, 2007); for example duplication events within the DNA sequence can give a false measure of variation amongst species (due to concerted evolution), and it is also possible to amplify and sequence ITS from fungal contaminants instead of the target plant species. Many

other plant barcode regions have been proposed (Hollingsworth, Graham, et al., 2011a, Dong, Xu, et al., 2015) and such reviews cover a wide range of examples and discuss the limitations of the technique. As cultivars are often of interspecific hybrid origin, DNA barcoding as originally proposed is only suitable to identify the seed parent, but not to distinguish one cultivar from another. The general principles of barcoding and the accompanying approaches to data analysis can be applied to a wider range of DNA regions and this could provide the resolution needed for cultivar identification if enough sequence data is generated.

1.5.5 Next Generation Sequencing (NGS)

The development of high-throughput sequencing over the last decade has provided researchers with vast amounts of data that has increased our knowledge of genomics, as well as biology. It is now possible to generate millions of short sequence reads (up to 700 bp) per sample, and use pipelines to assemble these reads together into long sequences (Unamba, Nag, et al., 2015). At an early stage, this can be challenging, particularly for non-model plants with no existing reference. The challenges of assembly are compounded for plants with large genomes, polyploid species or genomes that have undergone substantial gene rearrangement (Li, Lin, et al., 2017). Some of these issues can be overcome by using long read sequencing, which can produce sequences over 10kb, to build scaffold sequences (Li and Harkess, 2018). However, the error rate of base calls is currently high for long read sequencing, approximately 10%, and remains expensive per sample, a hybrid method of short and long reads sequencing is therefore favourable (Wang, Schalamun, et al., 2018). By comparison, traditional Sanger sequencing (referred to as DNA sequencing until the 2010s) can produce up to 1,400 bases in a single reaction, but for a single fragment, and by comparison the cost remains high. This highlights the increase in access to vast amounts of genomic data.

Numerous NGS methods have been developed, differing in the type of starting genetic material (RNA or DNA), or the method used to prepare the genome for sequencing (as reviewed by; Dodsworth, 2015, Goodwin, McPherson, et al., 2016). The combination of short-read and long-read sequencing can even allow the assembly of complex genomes (Nowoshilow, Schloissnig, et al., 2018). Next-generation sequencing is increasingly contributing to plant systematics, as researchers are able to transition from sequencing a handful of genes, to sequencing hundreds of genes or even sequencing whole genomes. Furthermore, as all of

NGS platforms sequence fragmented DNA in parallel reactions, sequencing libraries can even be constructed from very low amounts, even nanograms of DNA (Ross, Barrett, et al., 2016). Even DNA degraded below the usable quality needed for traditional sequencing methods can be used, this potentially allows small fragments of plant material from herbarium specimens to be sequenced (Bakker, Lei, et al., 2016).

Utilising NGS has had a profound effect upon marker development and phylogenetics (Uribe-Convers, Settles, et al., 2016) which will undoubtedly lead to a fundamental effect of DNA based cultivar identification. NGS provides the means to develop markers inexpensively and rapidly for further taxonomic study. Using a multiple genome approach (nuclear, plastid, mitochondrial) and sequencing several closely related taxa can improve the quality of data gathered and subsequent marker application. For example, over 9500 microsatellite primers were identified from seven peony genomes (Gilmore, Bassil, et al., 2013) of which 12 microsatellites were selected and applied to fingerprint 93 peony cultivars. The rapid advancement of sequencing technologies has caused computational challenges, as pipelines are currently not widely available to assemble whole genomes for plant species with large, complex genomes (Marschall, Marz, et al., 2016, Kyriakidou, Tai, et al., 2018).

Regardless of the system used, the advance of NGS will make discrimination of organisms that are closely related, have complex relationships or/and a hybrid origin both possible and practicable in the very near future. At present, the barrier to wide use remains the current cost per sample and the development of computational systems for whole genome assembly for large, complex genomes.

1.6 Daffodils in Cultivation

Daffodils are a popular spring plant and synonymous with many people in the UK as the beginning of spring, whilst in North America the Paperwhite daffodil, *N. papyraceus*, is a popular Christmas gift. The plants are an important part of the horticultural trade, being sold as both cut flowers and bulbs. A recent survey revealed that daffodils are the fourth most popular flower in the UK behind roses, lilies, and tulips (Avis-Riordan, 2017) and within the top five bestselling cut flowers within the UK as of 2006 (Flowers & Plants Association, 2006). The UK economic value of daffodils was valued at £45m per annum in 2015 (MarketInsider, 2015),

supporting a vast network of exports largely to Europe and North America and imports from the Netherlands (Hanks, 2002a).

Daffodils have been in cultivation for centuries, and were introduced to the UK by the 16th Century (Coates, 1956). It has been recorded that wild populations of daffodils in Spain and Portugal were exploited since the 11th Century as garden plants across Europe (Rivera Nuñez, Obón De Castro, et al., 2003). However, intensive breeding programmes in the last century in particular have given rise to over 30,000 registered cultivars (Kington, 2008, Daffseek, 2018). Of around 30,000 registered cultivars approximately 2,000 cultivated varieties are currently available in the UK (Cubey, 2017). The Royal Horticultural Society is the International Cultivar Registration Authority for daffodils, although the origin of many cultivars pre-dates the original register in 1907. Cultivar synonymy is widely reported in the register (Kington, 2008), and mis-labelled and mis-sold bulb stocks are widely known.

Daffodil cultivars are classified into 12 horticultural Divisions based upon morphological characters (Table 1-5). The purpose of the horticultural Divisions is to classify the variation of cultivars being displayed at horticultural shows and for garden purposes (Kington, 2008). Division 12 is for daffodil cultivars that cannot be grouped into any other horticultural Division, and all naturally occurring taxa are grouped within Division 13 including hybrids with a botanical name.

Table 1-5 Horticultural classification of cultivated daffodils. Descriptions according to Kington (2008). Each unit of the scale bar equals 1cm.

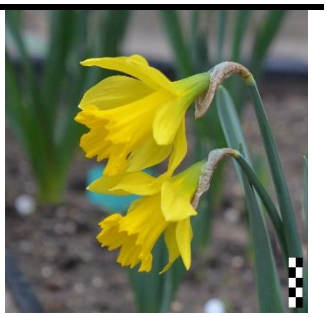
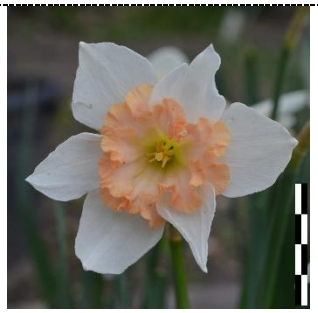
Division 1	Trumpet daffodil cultivars	One flower to a stem; corona ("trumpet") as long as, or longer than the perianth segments ("petals")	
Division 2	Large-cupped daffodil cultivars	One flower to a stem; corona ("cup") more than one-third, but less than equal to the length of the perianth segments ("petals")	
Division 3	Small-cupped daffodil cultivars	One flower to a stem; corona ("cup") not more than one-third the length of the perianth segments ("petals")	
Division 4	Double daffodil cultivars	One or more flowers to a stem, with doubling of the perianth segments or the corona or both	
Division 5	Triandrus daffodil cultivars	Characteristics of *N. triandrus* clearly evident: usually two or more pendent flowers to a stem; perianth segments reflexed	

Table 1-5 (cont.) Horticultural classification of cultivated daffodils. Descriptions according to Kington (2008). Each unit of the scale bar equals 1cm.

Division 6	Cyclamineus daffodil cultivars	Characteristics of *N. cyclamineus* clearly evident: one flower to a stem; perianth segments significantly reflexed; flower at an acute angle to the stem, with a very short pedicel ("neck")	
Division 7	Jonquilla & Apodanthus daffodil cultivars	Characteristics of Sections *Jonquilla* or *Apodanthi* clearly evident: one to five (rarely eight) flowers to a stem; perianth segments spreading or reflexed; corona cup-shaped, funnel-shaped or flared, usually wider than long; flowers usually fragrant	
Division 8	Tazetta daffodil cultivars	Characteristics of Section *Tazettae* clearly evident: usually three to twenty flowers to a stout stem; perianth segments spreading not reflexed; flowers usually fragrant	
Division 9	Poeticus daffodil cultivars	Characteristics of *N. poeticus* and related species clearly evident; perianth segments pure white; corona very short or disc-shaped, not more than one-fifth the length of the perianth segments; corona usually with a green and/or yellow centre and red rim, but sometimes wholly or partly of other colours; anthers usually set at two distinct levels; flowers fragrant	

Table 1-5 (cont.) Horticultural classification of cultivated daffodils. Descriptions according to Kington (2008). Each unit of the scale bar equals 1cm.

Division 10	Bulbocodium daffodil cultivars	Characteristics of Section Bulbocodium clearly evident: usually one flower to a stem; perianth segments insignificant compared with the dominant corona; anthers dorsifixed (i.e. attached more or less centrally to the filament); filament and style usually curved	
Division 11A	Split corona daffodil cultivars	Collar Daffodils Split-corona daffodils with the corona segments opposite the perianth segments; the corona segments usually in two whorls of three	
Division 11B	Split corona daffodil cultivars	Papillon Daffodils Split-corona daffodils with the corona segments alternate to the perianth segments; the corona segments usually in a single whorl of six	
Division 12	Other daffodil cultivars	Daffodil cultivars which do not fit the definition of any other Division	
Division 13	Daffodils distinguished solely by botanical name		

1.7 Aims

The aims of this thesis include:

1. To explore the natural level variation across populations of the *Narcissus minor* group (subsection *Pseudonarcissi*), to assess the implications for the species boundaries.
2. To assemble a reference plastid genome for *Narcissus.*
3. To assemble further plastid genomes for *Narcissus* and mine variable regions to build an initial genomic phylogeny for the genus.
4. To assess the utility of published microsatellite markers (Könyves, David, et al., 2016) in identifying daffodil cultivars from all horticultural Divisions.

Chapter 2 Wild populations of the *Narcissus minor* group

2.1 Introduction

The species limits and genetic differentiation of four north Iberian trumpet daffodils (*Narcissus* subsect. *Pseudonarcissi*) are investigated in this chapter. Three of these: *N. minor* L., *N. asturiensis* (Jord.) Pugsley, and *N. jacetanus* Fern. Casas are often referred to as the '*N. minor* group' (Pugsley, 1933). A further species *N. cyclamineus* DC. has been grouped with them on cytological (Fernandes, 1975), and phylogenetic grounds (Marques, Aguilar, et al., 2017).

Daffodils are popular spring flowers however, the taxonomy of the species remains uncertain; the taxonomy underpins their ties to horticulture through breeding programmes, and their conservation. Over the past half a century, there have been many taxonomic studies of the genus *Narcissus*, using morphology, cytology, and molecular data for classification (Fernandes, 1975, Webb, 1980, Zonneveld, 2008, Aedo, 2013, Marques, Aguilar, et al., 2017, Könyves, David, et al., 2019). This has led to a wide range of taxa being recognised, from 16 (Baker, 1875) to 160 (Haworth, 1831). However, previous studies have been confined to geographic regions (Webb, 1980, Aedo, 2013), stemmed from horticultural interests (Pugsley, 1933), or limited sampling due to the scope of the research (Barrett and Harder, 2004, Pérez-Barrales, Vargas, et al., 2004, Hodgins and Barrett, 2007). The last comprehensive monographic revision of the genus was completed in 1875 (Baker, 1875) however, since this time many new species have been described, and several species are known to have originated from cultivation (Wylie, 1952). This is highlighted by the probable garden origin of the type specimen of *N. pseudonarcissus* (LINN412.2), as it has a double flower. Despite the taxonomic uncertainty, there is consensus to split the genus into two subgenera: *Narcissus* and *Hermione*. These subgenera are further divided into 10 sections based upon morphological grounds (Fernandes, 1968, Aedo, 2013), although molecular data provides evidence for 14 sections (Marques, Aguilar, et al., 2017).

One of the most speciose and widespread groups in *Narcissus* are the trumpet daffodils (subsection *Pseudonarcissi* DC.), with a range from 13 (Webb, 1980) to 37 (Mathew, 2002) recognised taxa. The number of species within subsection *Pseudonarcissi* varies depending on whether taxa are recognised at the level of species, subspecies or variety (Blanchard, 1990), based upon the breadth of the species concept used by the author to recognise variation. The Iberian Peninsula is the centre of diversity for trumpet daffodils however, the distribution

extends north into France, beyond this range it is unknown whether the plants are genuinely native (Blanchard, 1990).

The subsection was formerly recognised as the genus *Ajax*, and subsequently as section *Pseudonarcissi* based upon morphological evidence (Salisbury, 1812). Species of trumpet daffodils have a corona at least as long as the tepals. However, recent molecular studies have revealed trumpet daffodils are a polyphyletic group (Graham and Barrett, 2004, Marques, Aguilar, et al., 2017, Könyves, David, et al., 2019). This led to splitting the trumpet daffodils found in Southern Spain from the section into section *Nevadensis*, based upon cytological and molecular grounds (Zonneveld, 2008, Marques, Aguilar, et al., 2017). Therefore, section *Pseudonarcissi* now includes all trumpet daffodils with a solitary flower. The most comprehensive molecular study of the genus to date, in terms of sampling and amount of data per sample, found that section *Pseudonarcissi* remains polyphyletic as section *Narcissus* was nested within the clade (Marques, Aguilar, et al., 2017). This led to the change in taxonomic rank to subsection *Pseudonarcissi*, within section *Narcissus*. Moreover, other sections were also found to be polyphyletic, sections *Tazettae*, *Jonquillae*, *Bulbocodii*. This suggests frequent hybridisation occurs within the genus, and subsequent gene flow amongst species prevents monophyletic groups being identified. This has even led to species groups being recognised, that were formerly identified at species level.

One such species group is the *N. minor* group and *N. cyclamineus* that forms a clade based upon data from mitochondrial and plastid genomes, and nuclear ribosomal DNA. The distinction between species within the *N. minor* group is complex, because the quantitative morphological characters used for identification have overlapping ranges (Table 2-1). The complexity of separating the species is compounded as species descriptions poorly characterise the plants especially for *N. jacetanus*, characters such as a larger flower, and considerably larger corona tube compared to *N. asturiensis* are used without quantification (Fernández Casas, 1984b). Thus, species such as *N. jacetanus* are distinguished by the use of characters that themselves are variable (Fernández Casas, 1984). Flora Iberica has not accepted *N. jacetanus* and regarded the species as a synonym of *N. minor* (Aedo, 2013). Whereas, Zonneveld (2008) recognised *N. jacetanus* based upon the DNA content, however to date the species has not been critically assessed. The wild origin of *N. minor* was originally unknown,

until being found in the Serra de Gerez in Portugal in the early 20th Century (Pugsley, 1933). However, the species was known in cultivation at least by the 16th century (Wylie, 1952). Accounts based upon morphological studies have accepted the species (Fernandes, 1968, Webb, 1980), the most recent account, Flora Iberica (Aedo, 2013), treats *N. asturiensis* as a subspecies of *N. minor*. Separating *N. asturiensis* from members of the *N. minor* group is complex, as the species varies extensively in the size of floral characters and overall plant stature (Blanchard, 1990).

Table 2-1 Morphological variation for species within the *N. minor* group, data taken from Webb (1980), Fernandez Casas (1984b), Blanchard (1990), and Zonneveld (2008).

Character	***N. asturiensis***	***N. minor***	***N. jacetanus***
Bulb size	15-20mm	20-30mm	Not recorded
Leaf size (L x W)	50-150mm x 2-6mm	80-200mm x 3-10mm	120mm x 9mm wide
Scape length	7-14cm	8-25cm	≤18cm
Spathe length	15-22mm	25-40mm	Not recorded
Flower	Solitary and drooping	Solitary, horizontal/drooping	Solitary, horizontal
Pedicel length	3-10mm	3-20mm	2-5 mm
Hypanthial tube (L x W)	8mm x 5-9mm	10mm x 9-15mm	12mm x 9mm
Perianth Segments (L x W)	7-14mm x 4mm	16-22mm	20mm x 8mm
Corona length	8-17mm	16-25mm	21mm
Corona margin	Margin somewhat lobed	Margin variously lobed	Deeply lobed
Colour	Medium yellow	Deep yellow	Yellow
Ploidy	Diploid	Diploid	Diploid

There is one key morphological character to separate species of the *N. minor* group from *N. cyclamineus*, the tepals of taxa within the *N. minor* group are not reflexed while those of *N. cyclamineus* are strongly so. Based upon morphological evidence, the recognition of *N. cyclamineus* has not been questioned however, molecular evidence has been unable to separate the species from the *N. minor* group. A phylogeny based upon plastid DNA found a polytomy between the species (Graham and Barrett, 2004), and nuclear ribosomal, plastid and mitochondrial DNA was unable to separate *N. cyclamineus* from *N. asturiensis* (Marques, Aguilar, et al., 2017).

The species within the study group have overlapping distributions (Figure 2-1); *N. jacetanus* is distributed to the east of the range of the *N. minor* group, *N. cyclamineus* to the west, however,

N. minor and *N. asturiensis* overlap across the north of the Iberian Peninsula. The species are found at a range of altitudes (0-2000m) and varying habitats: *N. cyclamineus* is found along river banks ranging from altitudes of 0-1000m, *N. asturiensis* can be found growing amongst *N. cyclamineus* and at altitudes up to 2000m. *Narcissus jacetanus* is found growing in calcareous soil at a range of altitudes from 500-2000m, similar to the habitat of *N. minor* (Lorda López, 2013).

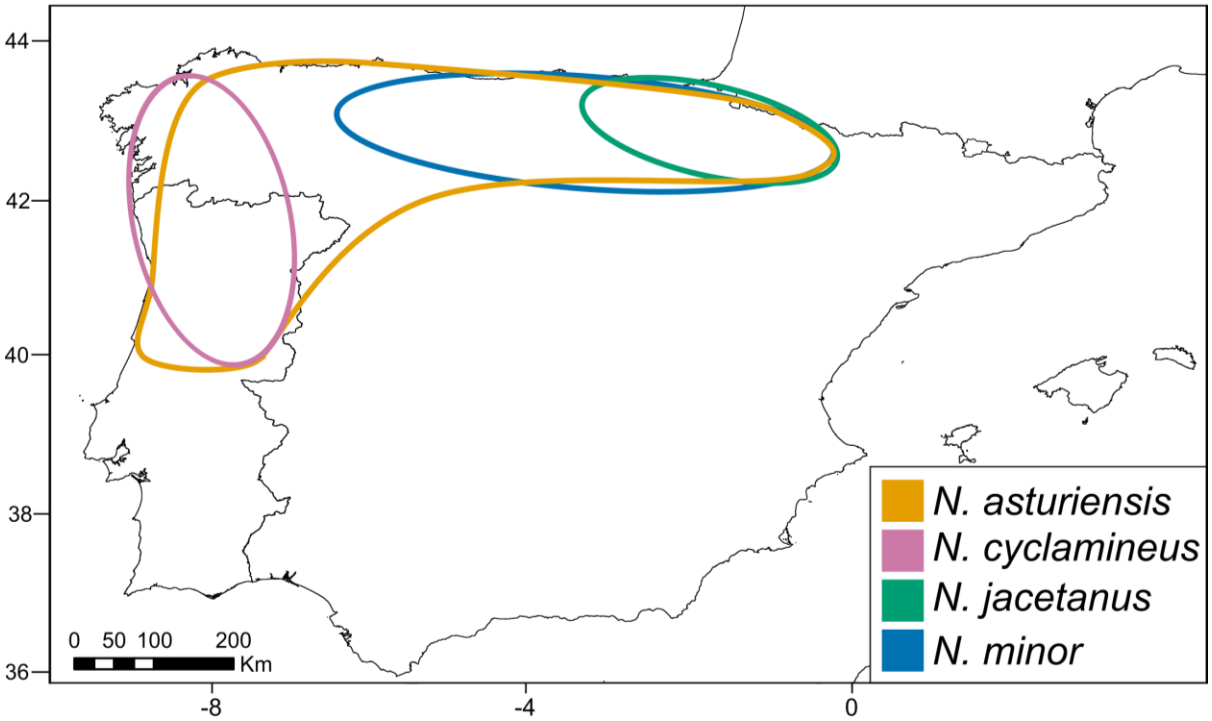


Figure 2-1 Distribution map of the four study taxa. Regions have been drawn from corrected GBIF occurrence data for four of the species (GBIF.org, 2018).

Habitats of the *N. minor* group are currently threatened by agriculture and forestry plantation (Lomba, Pellissier, et al., 2010) and the plants have been illegally collected for commercial trade (Caldas and Moreno Saiz, 2011). Two of the species *N. cyclamineus* and *N. asturiensis* are IUCN red listed (IUCN, 2018), and therefore covered by legal protection (EU Directive 92/43/EEC). Both species are classified as Least Concern, however *N. cyclamineus* has no assessment data. The IUCN taxonomic treatment is based upon Flora Iberica (Aedo, 2013), although *N. minor* is classified as a species it has not been assessed, whereas *N. jacetanus* is classified as a synonym of *N. minor*. Taxonomic study is required for providing a predictable nomenclature, which is essential for species conservation (Thomson, Pyle, et al., 2018).

Genera with similar taxonomic complexities arising from prevalent hybridisation, have previously been studied using a combination of DNA fingerprinting alongside DNA sequencing (Jang, Müllner, et al., 2005). For example, microsatellites were used to observe species boundaries in *Quercus* to identify parentage of naturally occurring hybrid taxa (Castillo-Mendoza, Salinas-Sánchez, et al., 2018). Microsatellites combined with plastid sequencing were also able to ascertain species boundaries in *Dioscorea* and identify the natural hybridisation to explaining intermediate morphotypes (Chaïr, Sardos, et al., 2016). Using multiple plant genomes in such studies can also identify phylogeographical patterns (Cutter, 2013).

Within *Narcissus,* microsatellites have previously been used alongside DNA sequencing to ascertain species boundaries within hoop-petticoat daffodils (Könyves, David, et al., 2019), to identify varieties of *Narcissus triandrus* (Hodgins and Barrett, 2007), and to examine population level variation of *N. papyraceus* (Simon-Porcar, Pico, et al., 2015). Inter-simple sequence repeats have also been used to elucidate the taxonomy of section *Nevadensis* taxa distributed in southern Spain (Jiménez, Sánchez-Gómez, et al., 2009). It is therefore a useful approach to explore the natural variation and geographic patterns of variation within this complex group. In this study, we use microsatellite markers alongside cpDNA sequencing to examine whether the patterns of molecular variation coincide with the taxonomy of the *Narcissus minor* group and closely related taxa. This study is limited to the three most common species within the *N. minor* group, to identify if the levels of variation reflect the increase in names, outlined in Chapter 1.

2.2 Objectives

The aim of this chapter is to identify the population level variation amongst three representatives of the *Narcissus minor* group and the closely related *N. cyclamineus* to compare whether the levels of variation reflect the current taxonomy. Thus, the objectives in this chapter are:

1. To identify maternally inherited plastid DNA (*matK*) variation amongst the study group taxa.
2. To examine the population level genetic structure of microsatellite genotyped populations.

2.3 Methods

2.3.1 Source material

Plant sampling aimed to cover the entire distribution range for the four species involved in this study. A total of 615 individuals from 56 populations were sampled (population data detailed in Appendix A). Leaf material from up to 15 individuals per population was collected, where possible using a haphazard strategy, and stored in silica gel. As the plants can reproduce clonally samples were collected at least two metres apart to prevent collecting the same ramet. Voucher specimens for each population have been deposited at the University of Reading Herbarium (**RNG**). Plant material was also sourced from collaborators' field collections, and specimens of a living collection of wild provenance plants (Appendix A). Genomic DNA was extracted from silica dried material using a modified CTAB protocol (Doyle and Doyle, 1987). DNA quality and quantity were observed using 0.8% agarose gel electrophoresis stained with GelRed® (Biotium, Fremont, US) using Hyperladder™ 1Kb (Bioline, London UK) as a reference. Concentration was also confirmed using NanoDrop Lite (ThermoFisher, Paisley UK). DNA was diluted into 5ng/µl aliquots for subsequent PCR reactions.

2.3.2 Microsatellite screening

Microsatellites were genotyped using primers and protocols described in Könyves, David & Culham (2016), shown in Table 2-2. PCR reactions were performed in 10µl volumes containing a final concentration of 1 × Biomix (Bioline, London UK), 0.2µM of each primer, and 10ng of DNA product. Cycling conditions were 94°C for 2 minutes, followed by 40 cycles of 94°C for 45s, 48-63°C for 30s, 72°C for 45s, and an extension phase of 72°C for 10 minutes. The PCR products were visualised using 2% (w/v) agarose gel electrophoresis, stained with GelRed® (Biotium, Fremont US), using Hyperladder™ 100bp (Bioline, London UK) as a reference. PCR reactions were performed as single reactions and PCR products mixed for multiplexed capillary electrophoresis, groupings shown in Table 2-2. Capillary electrophoresis of the PCR products was completed by Source BioScience (Nottingham, UK) using an ABI 3730xl DNA Analyzer. Electropherograms were sized using in Geneious 11.0.3 microsatellite plugin 1.4.4 (Kearse, Moir, et al., 2012). Scoring was completed according to the MANUAL 8 protocol (Pfeiffer, Roschanski, et al., 2011), to prevent overestimation or incorrect peak calling. The MANUAL protocol aims to score each peak in the electropherograms from 1-10, 1 being a putative

stutter peak, and 10 a high peak at least 85-90% of the height of the main peak. All peaks given a score below 8 are discarded to prevent overestimating the variation of loci, hence MANUAL 8 protocol.

Table 2-2 Details of the microsatellites used in this study, information gathered from Könyves, David, et al. (2016).

Locus	Primer sequence	Repeat motif	Allele size range	T_a(°C)	Dye	Multiplex Group
NSB14	F: TGTGTAAGCATACTAACGTTTCG R: AAAAGAGCACCAAGGATGAA	$(ATGT)_{14}$	147-221	48	6-FAM	1
NSB113	F: TTGTGATAAATAAAGGTGCAACTCA R: CATTGCCCGTGATAAGCTCT	(AGT)6	78-123	54	6-FAM	2
NSB122	F: CAAAGTGTTTGTGAATTGCTTC R: GCAATGAGGAGCTTATGAATTAAC	$(AC)_9$	169-198	59	PET	2
NSB143	F: CTGTTTCTTTGTTCTGCACATT R: TCCCAAAATTGCTTCTGAGC	$(GAA)_7$	244-269	59	NED	3
NSB182	F: TTGTATTATACGTTGTTCTGGGGA R: GAGATGCTGACACGCAAACT	$(AC)_{12}$	115-121	58	VIC	1
NSB232	F: CTCCACTTTGGTTGAATCCC R: GACTACCTCCTATTCTAAATGCCA	$(CT)_9$	110-118	63	VIC	2
NSB253	F: GAGGATTACTGTAGCCAATTCCA R: GGACTACAAGATGGCTTCCA	$(GTT)_{11}$	100-140	56	NED	2
NSB263	F: CGAAGGAGGAGTCTTGGAAA R: GAGCAAACTCCTGGCTGAAG	$(GAA)_{11}$	94-131	60	VIC	3
NSB272	F: GGTTCTGCCGATGGACTAAT R: TTATCACATCCAACGGTTTGC	$(CT)_9$	122-146	60	NED	1

For each population, diversity estimators were calculated in GenAlEx 6.5 (Peakall and Smouse, 2012): the total (A) and average number of alleles (H'), the number of private alleles (A_p), the observed heterozygosity (H_o), the proportion of null genotypes (F_{g0}), the number of multilocus genotypes (G), and the complement of Simpson's diversity index (D) were all calculated.

Discriminant analysis of principal components (DAPC) was used to identify clusters in the microsatellite data using the adegenet 2.1.1 (Jombart, 2008) package in *R*. Prior groups were defined using sequential k-means clustering algorithm, increasing the value of k and comparing clustering scenarios to Bayesian Information Criterion values (Jombart and Collins, 2015). The cumulative variance explained by the PCA eigenvalues was observed and PCs were retained to represent 95% of the total variance. The optimal number of clusters was identified as k=4. All of the eigenvalues were retained, as the number of clusters was small. DAPC analysis was repeated with the 3 most populous clusters (k=3) to observe if there was any further genetic structure.

A mean pairwise distance matrix for populations was calculated, using the data from all populations, and the populations in the 3 most populous clusters, in the *R* package ade4 ver. 1.7-13 (Dray and Dufour, 2007). The correlation between genetic distance and geographic distance was then estimated using a Mantel Test. The Mantel test was calculated using the package fossil ver 0.3.7 (Vavrek, 2011) in *R*.

2.3.3 DNA sequencing

Variation of the *matK* region was surveyed using universal plastid DNA primers *matK*2.1 and *matK*5 (Ford, Ayres, et al., 2009) to amplify part of the *matK* gene for all 615 individuals. PCR reactions were performed in 25µl volumes containing a final concentration of 1 × Biomix (Bioline, London UK), 0.35µM forward and reverse primers, 1mg/ml BSA, 4% v/v DMSO, and 10ng/µl DNA template. Cycling conditions were 94°C for 2 minutes, followed by 30 cycles of 94°C for 30 seconds, 48°C for 30 seconds and 72°C for 1 minute with a final extension period of 72°C for 7 minutes. A section of two further genes *ndhF* and *ycf1* were amplified for one representative of each population. Primers *ndhF* 745F and *ndhF* 2110R (Terry, Brown, et al., 1997) were used to amplify partial *ndhF* gene and *ycf1*bF and *ycf1*bR (Dong, Xu, et al., 2015) were used to amplify partial *ycf1* gene. PCR of *ndhF* reactions were performed as in *matK* reactions. Cycling conditions for *ndhF* were 94°C for 2 minutes, followed by 30 cycles of 94°C for 30 seconds, 52°C for 60 seconds and 72°C for 2 minutes with a final extension period of 72°C for 7 minutes. PCR amplification of partial *ycf1* gene were performed as for *matK*, the cycling conditions were also identical. Purified PCR products were sequenced in both directions by Eurofins Genomics (Ebersberg, DE).

Consensus sequences were assembled in Seqman Pro version 13.0.2.422 (DNAStar, Madison US). Sequences for *matK* were combined with the data from (Könyves, David, et al., 2019) in order to provide the context of the study group to other sections of *Narcissus,* included sequences are shown in Appendix A. Gene sequences of *matK*, *ndhF* and *ycf1* for two outgroup species: *Pancratium maritimum* L. and *Lapiedra martinezii* Lag. were extracted from data in Chapter 4 of this thesis to provide outgroups for phylogenetic analysis. Sequences were aligned in AliView 1.23 (Larsson, 2014) using the MUSCLE 3.8.425 algorithm (Edgar, 2004). A statistical parsimony network was constructed from *matK* dataset using TCS 1.21 (Clement, Posada, et

al., 2000), to explore the variation across the genus, and relationship amongst sampled populations.

Phylogenic trees were constructed using Bayesian Inference (BI) analysis, in MrBayes 3.2.6 (Ronquist, Teslenko, et al., 2012). The best-fit model for the dataset was selected using hierarchical likelihood ratio tests and Akaike information criterion, implemented in MrModeltest 2.4 (Nylander, 2004). All BI trees were conducted with two separate runs with four chains. Both combined datasets ((*matK* and *ndhF*) and (*matK*, *ndhF*, and *ycf1*)) were run for 2,000,000 generations, sampling every 1,000 generations. Burn-in was identified by assessing convergence of each run in Tracer 1.7 (Rambaut, Drummond, et al., 2018). Trees from the first 15% sampled generations were discarded. Combinable component consensus trees were generated in BayesTrees v1.3 (Pagel, Meade, et al., 2004), and used as they show the best supported clades.

2.4 Results

2.4.1 Microsatellite diversity

A total of 69 alleles were scored from the amplified loci in 615 individuals from 56 populations across the distribution range. The number of alleles per microsatellite marker ranged from 5 alleles at both locus NSB143 and NSB182, and 12 alleles at locus NSB272. A summary of the genetic variation within each of the sampled populations is shown in Table 2-3. The number of alleles (A) per population ranged from 14 and 28. Populations with a larger number of sampled individuals did not consistently have an increased number of alleles. For example, population MON: N=9, A=22, compared with BDL: N=15, A=21. The proportion of heterozygotes ranged from 0.16 (MDS) to 0.47 (P2 and SCO). The number of null alleles (F_{g0}) was consistently low. The mean F_{st} value was 0.29 (SD ± 0.064), indicative of low population differentiation.

Table 2-3 Summary statistics for the genetic diversity of sampled populations. * = species name under which sample was collected; N=the number of individuals, A=the total number of different alleles across all loci; H'=the mean number of different alleles across all loci; A_p=the number of private alleles across all loci; H_O=proportion of observed heterozygotes per loci; F_{g0}=proportion of null genotypes carried by each individual averaged across all loci; G=the number of multilocus genotypes; D= the complement of Simpson's diversity. Haplotype and DAPC cluster identified for each population is included: M=*matK* haplotype, K=DAPC cluster.

Pop.	Species*	N	A	H'	A_p	H_O	F_{g0}	G	D	M	K
AGA	*N. cyclamineus*	9	24	2.67	0	0.43	0	9	0.89	H5	2, 3
AGAAST	*N. asturiensis*	3	18	2	0	0.41	0	3	0.67	H5	2
ALT	*N. asturiensis*	10	21	2.33	1	0.31	0.002	10	0.9	H7	2, 3
ARA	*N. asturiensis*	10	24	2.67	1	0.31	0	10	0.9	H9	3, 4
Aralar	*N. minor*	16	25	2.78	0	0.31	0	16	0.94	H7	1, 3
Arg	*N. asturiensis*	12	24	2.67	0	0.28	0	12	0.92	H7	3
BDL	*N. jacetanus*	15	21	2.33	0	0.45	0.002	14	0.92	H5	1, 3
BEM	*N. cyclamineus*	5	18	2	0	0.38	0.002	5	0.8	H5	2
BRA	*N. jacetanus*	10	28	3.11	4	0.41	0	10	0.9	H7, H8	2, 3
CAN	*N. cyclamineus*	10	20	2.22	0	0.32	0.002	10	0.9	H5	2
CDB	*N. asturiensis*	10	23	2.56	0	0.31	0.01	10	0.9	H7	2, 3
CEB	*N. asturiensis*	10	22	2.44	0	0.31	0.007	10	0.9	H5	2, 3
CUE	*N. cyclamineus*	12	24	2.67	0	0.32	0	12	0.92	H5	2, 3
DEZ	*N. cyclamineus*	10	23	2.56	0	0.33	0	10	0.9	H5	3
Entzia	*N. asturiensis*	15	23	2.56	0	0.33	0	15	0.93	H7	1
FRI	*N. cyclamineus*	14	28	3.11	0	0.25	0	14	0.93	H5	3
GOL	*N. asturiensis*	10	26	2.89	4	0.28	0	10	0.9	H7	2, 3
ISI	*N. asturiensis*	10	25	2.78	0	0.36	0.002	10	0.9	H7	2, 3
LDS	*N. asturiensis*	10	25	2.78	0	0.28	0	10	0.9	H7, H9	3, 4
LER	*N. cyclamineus*	10	22	2.44	0	0.34	0.002	10	0.9	H5	2, 3
LUR	*N. asturiensis*	10	25	2.78	0	0.3	0	10	0.9	H7	2, 3
MDS	*N. minor*	11	24	2.67	0	0.16	0	10	0.89	H9	3, 4
MON	*N. cyclamineus*	9	22	2.44	0	0.37	0	9	0.89	H5	3
P2	*N. cyclamineus*	20	27	3	0	0.47	0.002	20	0.95	-	3
P3	*N. cyclamineus*	20	27	3	0	0.34	0	19	0.95	-	3
P4	*N. cyclamineus*	20	28	3.11	0	0.32	0.01	20	0.95	-	2, 3
P5	*N. cyclamineus*	20	24	2.67	0	0.36	0	20	0.95	-	3
P6	*N. cyclamineus*	20	25	2.78	1	0.3	0.007	20	0.95	-	2, 3
PAC	*N. minor*	10	21	2.33	0	0.3	0.002	10	0.9	H5	2, 3
PBL	*N. asturiensis*	10	20	2.22	0	0.3	0	10	0.9	H7	1, 2, 3
PDF	*N. asturiensis*	15	20	2.22	0	0.28	0	14	0.92	H7	1
PdH	*N. asturiensis*	11	24	2.67	1	0.27	0	11	0.91	H7	1, 3
PdLi	*N. jacetanus*	15	24	2.67	1	0.31	0	14	0.92	H7	1
PDM	*N. asturiensis*	10	23	2.56	0	0.3	0	10	0.9	H7	2, 3
PDP	*N. jacetanus*	10	27	3	3	0.39	0	10	0.9	H7	2, 3
Per	*N. asturiensis*	15	23	2.56	0	0.28	0	15	0.93	H7	2, 3

Table 2-3 (cont.) Summary statistics for the genetic diversity of sampled populations. * = species name under which sample was collected; N=the number of individuals, A=the total number of different alleles across all loci; H'=the mean number of different alleles across all loci; A_p=the number of private alleles across all loci; H_O=proportion of observed heterozygotes per loci; F_{g0}=proportion of null genotypes carried by each individual averaged across all loci; G=the number of multilocus genotypes; D= the complement of Simpson's diversity. Haplotype and DAPC cluster identified for each population is included: M=*matK* haplotype, K=DAPC cluster.

Pop.	**Species**	**N**	**A**	**H'**	**A_p**	**H_O**	**F_{g0}**	**G**	**D**	**M**	**K**
PMA	*N. asturiensis*	11	22	2.44	0	0.37	0.002	10	0.89	H7	1
PON	*N. asturiensis*	11	26	2.89	0	0.33	0	11	0.91	H6	1, 3
POV	*N. asturiensis*	10	21	2.33	0	0.37	0.002	9	0.88	H7	3
RM	*N. asturiensis*	4	14	1.56	0	0.36	0	4	0.75	H5	2, 3
ROZ	*N. asturiensis*	10	27	3	0	0.4	0.003	10	0.9	H7	2, 3
SAN	*N. jacetanus*	10	22	2.44	0	0.5	0.002	10	0.9	H7	2, 3
Sar	*N. asturiensis*	10	21	2.33	0	0.33	0	9	0.88	H7	1, 3
SCO	*N. cyclamineus*	6	20	2.22	0	0.47	0.002	6	0.83	H5	2, 3
SDC	*N. cyclamineus*	3	16	1.78	3	0.44	0	3	0.67	H5	3
SDP	*N. minor*	10	25	2.78	0	0.46	0	10	0.9	H9	4
SEN	*N. asturiensis*	10	24	2.67	0	0.34	0	10	0.9	H7	2, 3
SIG	*N. cyclamineus*	11	25	2.78	2	0.36	0.002	10	0.89	H5	2, 3, 4
SJM	*N. cyclamineus*	10	18	2	0	0.27	0	10	0.9	H7	2, 3
SM	*N. asturiensis*	15	24	2.67	0	0.42	0.003	15	0.93	H7	1, 3
TEB	*N. cyclamineus*	7	18	2	2	0.3	0	7	0.86	H5	2
VDA	*N. jacetanus*	15	19	2.11	0	0.29	0.005	13	0.92	H7	1, 3
VM	*N. asturiensis*	15	22	2.44	0	0.18	0.002	13	0.91	H7	1, 3
Mean		11.42	22.87	2.54	0.42	0.34	0.001	11.17	0.89	-	-
std (±)		4.03	3.11	0.35	0.99	0.07	0.002	3.92	0.06	-	-

The discriminant analysis separated the entire microsatellite data of 615 individuals into four clusters (

Figure 2-2). A threshold of 50% membership probability was used to assign individuals to a cluster with most samples being assigned to only one cluster; few individuals have a membership probability from more than one cluster. The first discriminant functions accounts for 69% of the variation, the second for 21% and third for 8%. Cluster 3 was the most populous in terms of samples, accounting for 54% of individuals; cluster 2 contained 21% of individuals, cluster 1 with 17% and finally cluster 4 containing only 8% of sampled individuals. Cluster 1 is comprised of three taxa: *N. asturiensis*, *N. jacetanus*, and *N. minor*. Clusters 2 and 3 both correspond to four taxa: *N. asturiensis*, *N. cyclamineus*, *N. jacetanus*, and *N. minor*. Cluster 4

corresponds to four populations of *N. minor*. The loadings for each discriminant function are shown in Figure 2-3; loci from NSB122, NSB253, and NSB272 consistently make the highest loadings in all three of the discriminant functions.

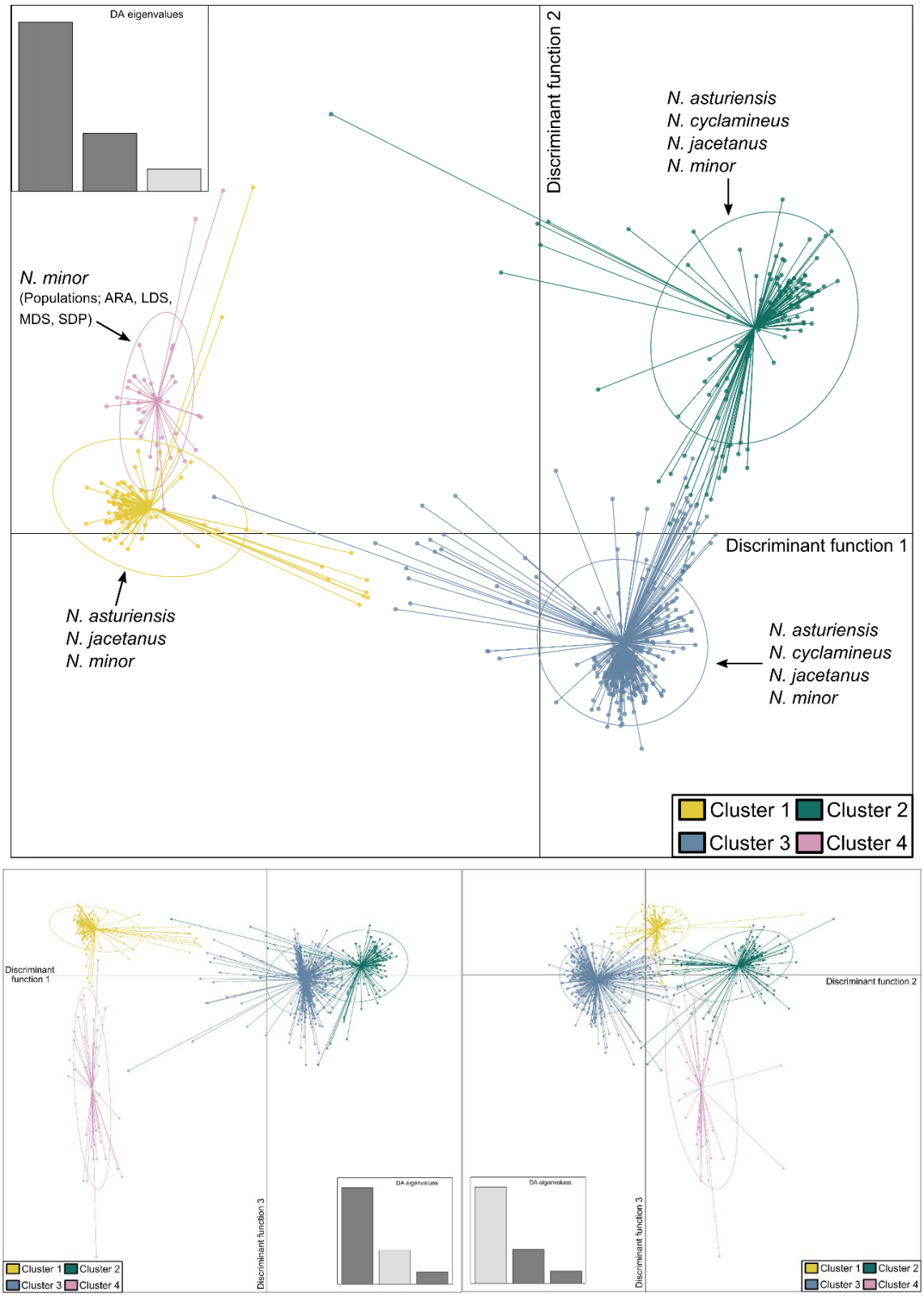


Figure 2-2 DAPC scatterplot based upon entire dataset of 615 genotyped individuals. Discriminant functions 1 and 2 (above), discriminant functions 1 and 3 (below left), and discriminant functions 2 and 3 (below right).

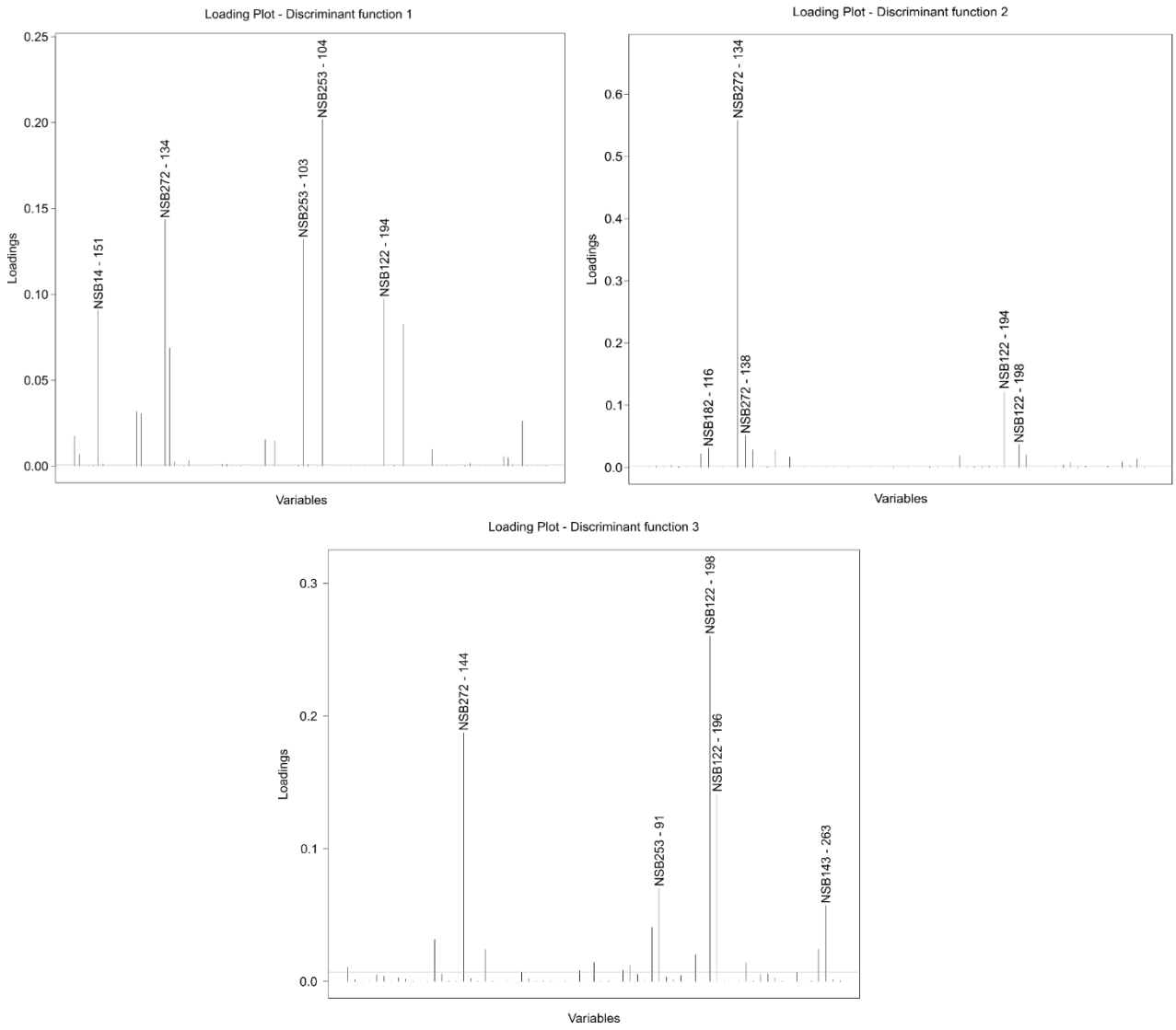


Figure 2-3 Loading plot for discriminant functions corresponding to Figure 2-2. The five loci with the highest loadings are labelled.

After re-analysing the dataset with populations in cluster 4 removed, there is overlap of the species found in each of the DAPC clusters (Figure 2-4). The first discriminant function accounts for 73% of the total variation, and discriminant function 2 accounts for 27%. Cluster 1 is comprised of two taxa: *N. asturiensis*, and *N. cyclamineus*. Clusters 2 and 3 each contain all the taxa from the *N. minor* group. The loadings for each discriminant function based upon the data set with cluster 4 populations removed are shown in Figure 2-5. Loci from NSB122 and NSB272 make the largest contribution to both discriminant functions.

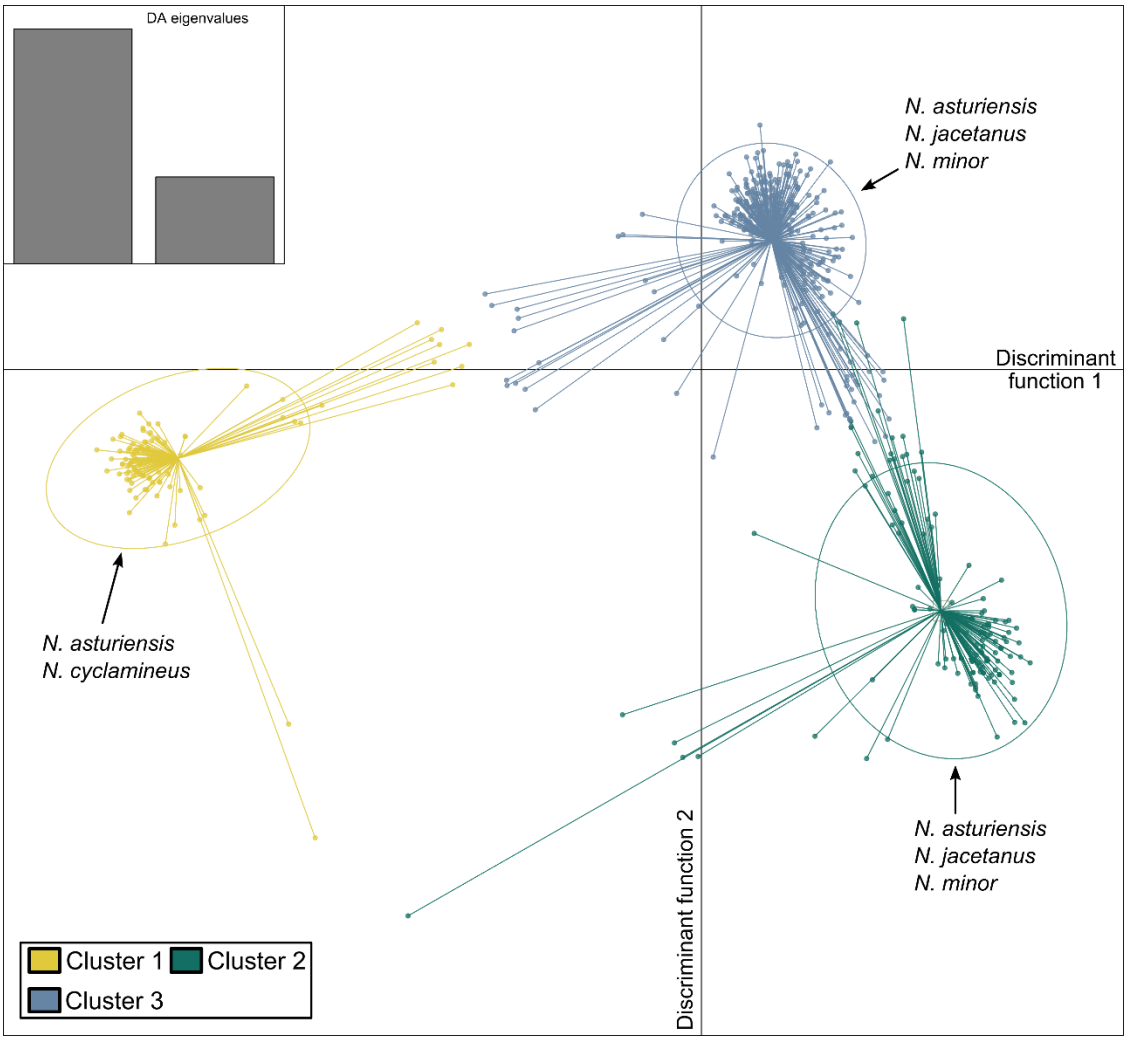
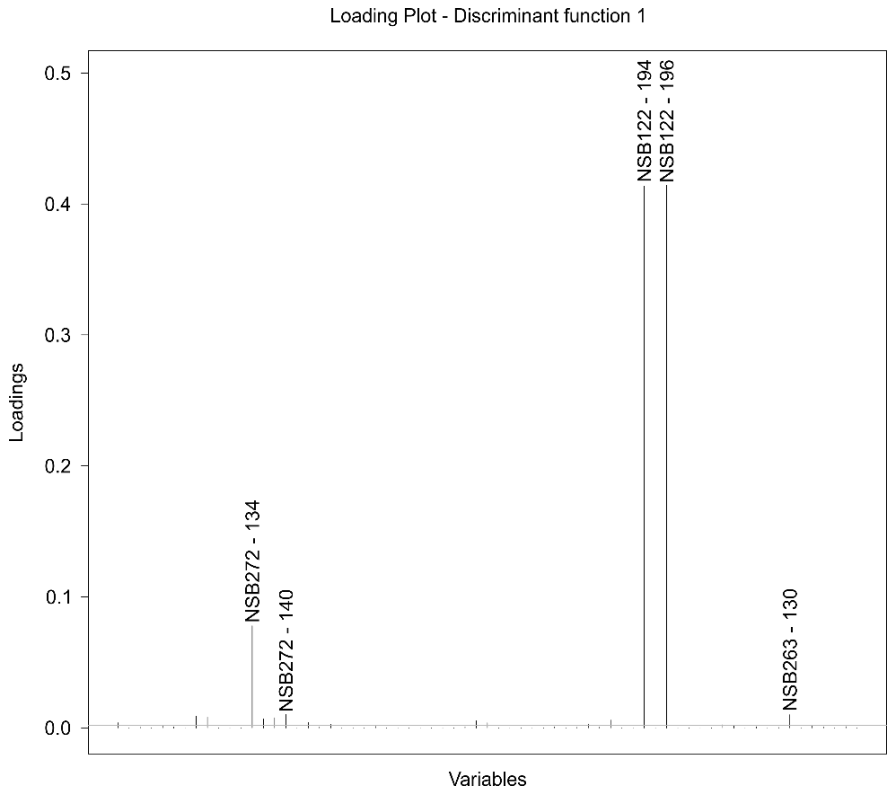


Figure 2-4 DAPC scatterplot based upon a subset of the whole dataset, cluster 4 populations from Figure 5 were removed, including 564 genotyped individuals. Discriminant function 1 and 2.



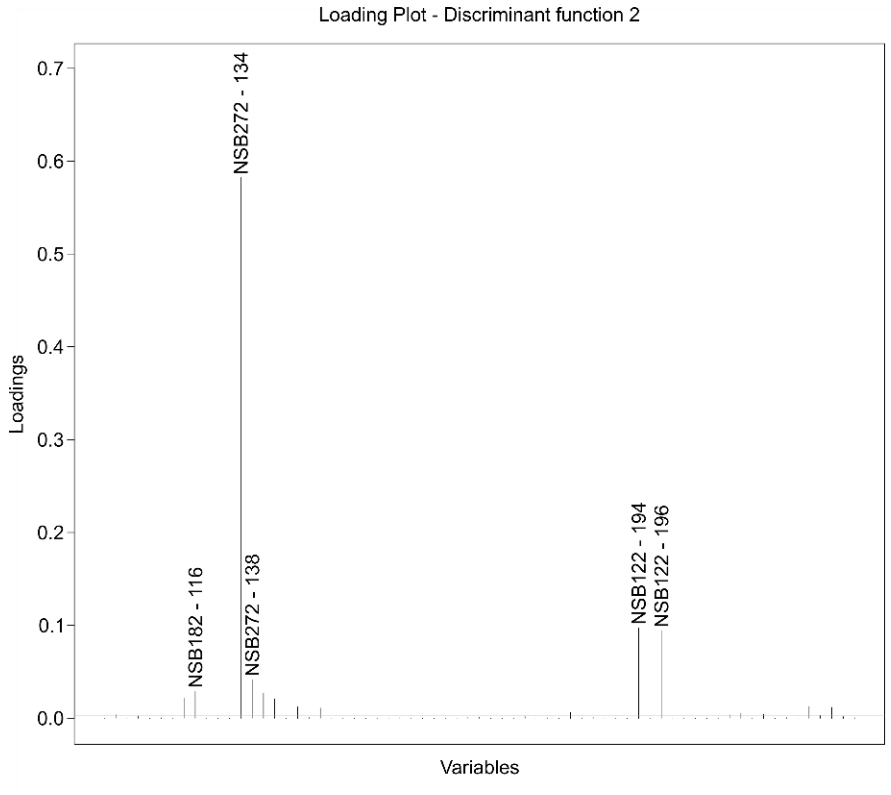


Figure 2-5 Loading plot for discriminant function 1 (above) and discriminant function 2 (below) corresponding to Figure 2-4. The five loci with the highest loadings are labelled.

The geographic structure of the DAPC analysis is shown in Figure 2-6. Cluster 3 is the most widespread, occurring throughout the *Narcissus minor* group's distribution range and within populations of *N. cyclamineus*. Cluster 2 occurs within the west of the distribution range until the east of the Cantabrian Mountains, replaced by cluster 1 moving east towards the Pyrenees. By comparison, cluster 4 is a grouping of four populations, corresponding to populations of *N. minor* sampled whilst not in flower. These populations are found to the east of the Leon Mountains, and south of the Cantabrian Mountains, besides one population (ARA) that occurs within the Cantabrian Mountain range. Two populations, SIG (*N. cyclamineus*) and PBL (*N. asturiensis*) belong to more than two clusters identified by the DAPC analysis; the remaining populations only correspond to a maximum of two clusters.

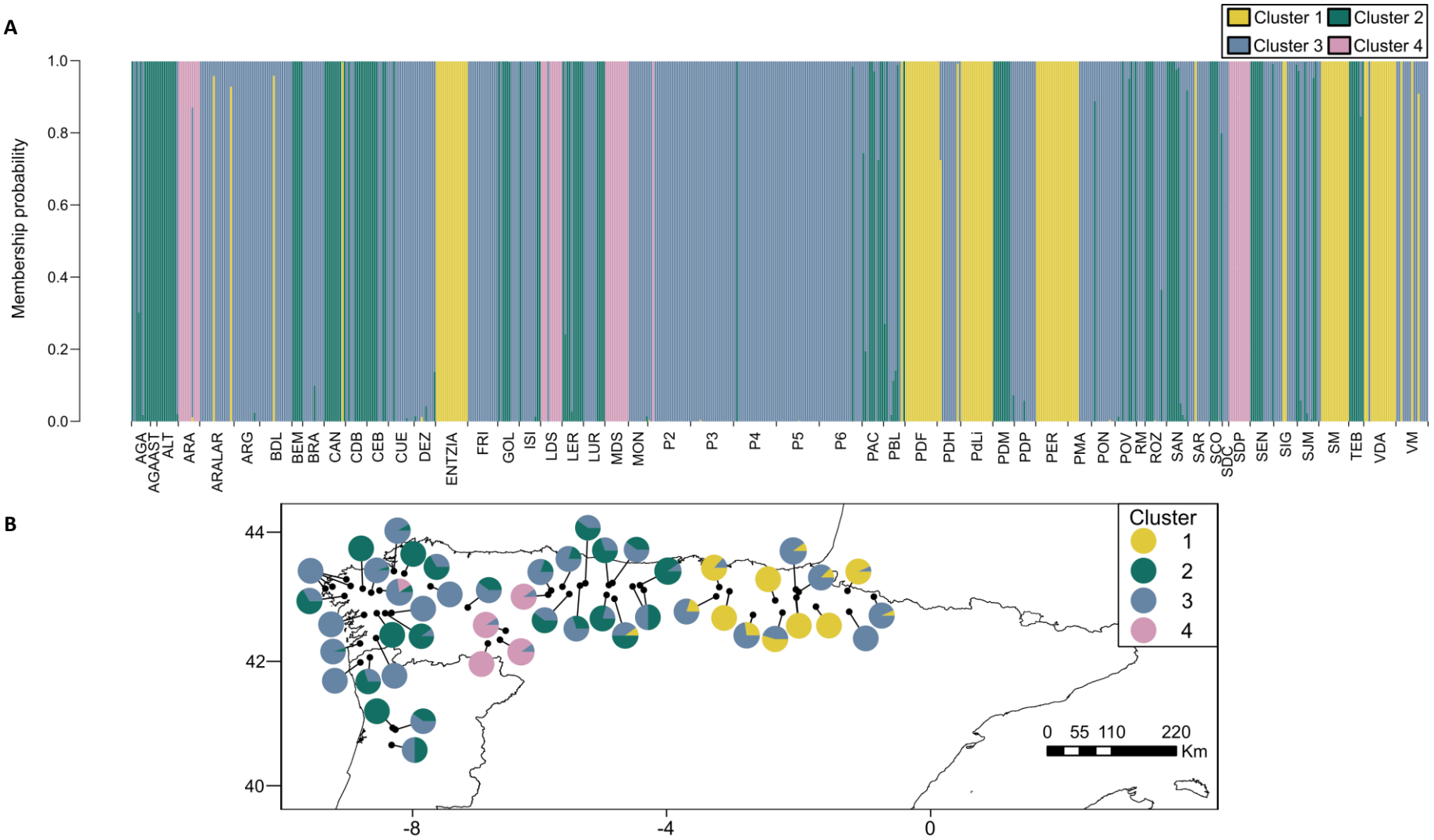


Figure 2-6 A; Bar plot of the DAPC analysis for 615 individuals of the study group. Based upon membership probability of each individual assigned to a cluster when K=4. B; Geographic distribution of the identified DAPC clusters.

2.4.2 Spatial autocorrelation

The Mantel test of the correlation between geographic distance and genetic distance found a positive yet weak correlation, rm = 0.181, P<0.05 (Figure 2-7). Change in geographic distance did correlate with genetic distance, which can be observed in the geographic population structure of the sampled populations (Figure 2-6). Genetic distance was limited as cluster 3 is found throughout the distribution range of the *N. minor* group. Removing populations corresponding to cluster 4 in the DAPC analysis found a stronger positive correlation, rm = 0.131, P<0.05 (Figure 2-8), compared with the Mantel Test with all populations.

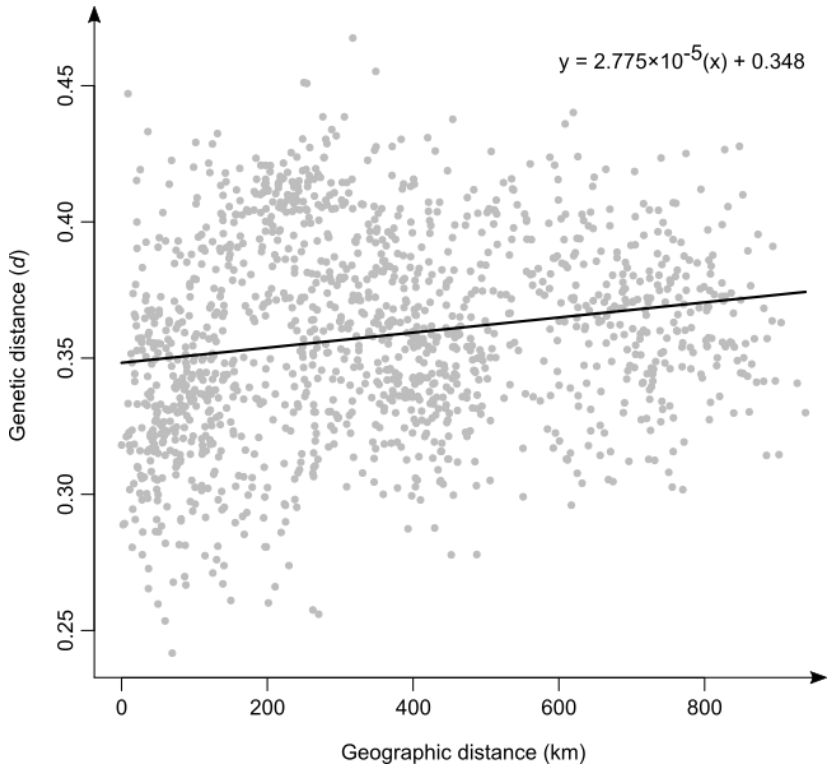


Figure 2-7 Mantel test of correlation between geographic distance and genetic distance, for entire dataset. The line of regression is overlaid. Genetic distance is calculated as; d = √1- Ssm, where Ssm is the simple matching coefficient of Sokal and Michener (1958).

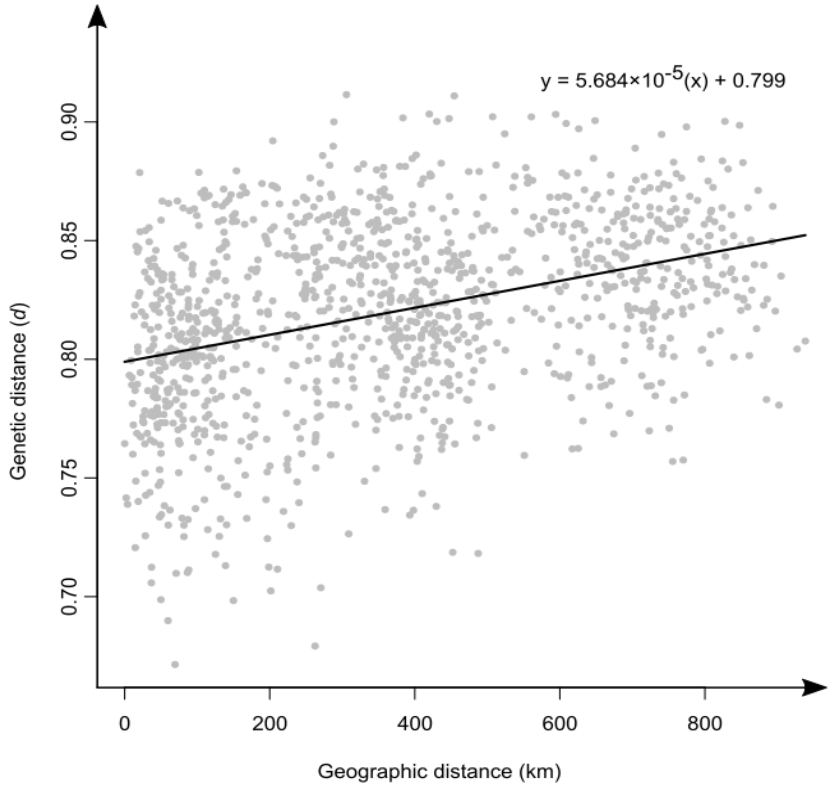


Figure 2-8 Mantel test of correlation between geographic distance and genetic distance, with cluster 4 populations removed. The line of regression overlaid. Genetic distance is calculated as; d = √1- Ssm, where Ssm is the simple matching coefficient of Sokal and Michener (1958).

2.4.3 Chloroplast DNA variation

The BI tree built from the *matK* and *ndhF* dataset (2,196bp) confirms *Narcissus* as monophyletic (Figure 2-9), with support for the subgeneric split of *Hermione* and *Narcissus*. Section *Narcissus* is polyphyletic, split into three clades including *N. poeticus* of subsection *Narcissus*, and species of subsection *Pseudonarcissi*. Taxa belonging to the *Narcissus minor* group and *N. nobilis* (subsect. *Pseudonarcissi*) form a well-supported clade, posterior probability of 1, although support within the clade is between 0.02 and 0.99. The groups of populations found with identical sequences are shown in Table 2-4. Cluster 4 of the DAPC analysis (

Figure 2-2) forms a well-supported clade with high support (0.99) within the clade of study group. The BI tree based upon three cpDNA regions (3,160bp), shows a similar topology (Figure 2-10). The *N. minor* group forms a clade, however *N. nobilis* cannot be separated from several taxa within the study group. There is overlap of the taxa, meaning species within the *N. minor* group and *N. cyclamineus* are not monophyletic, although as a group they well supported. The

minimum posterior probability is low in both BI trees, with 0.02 as the minimum probability for a node.

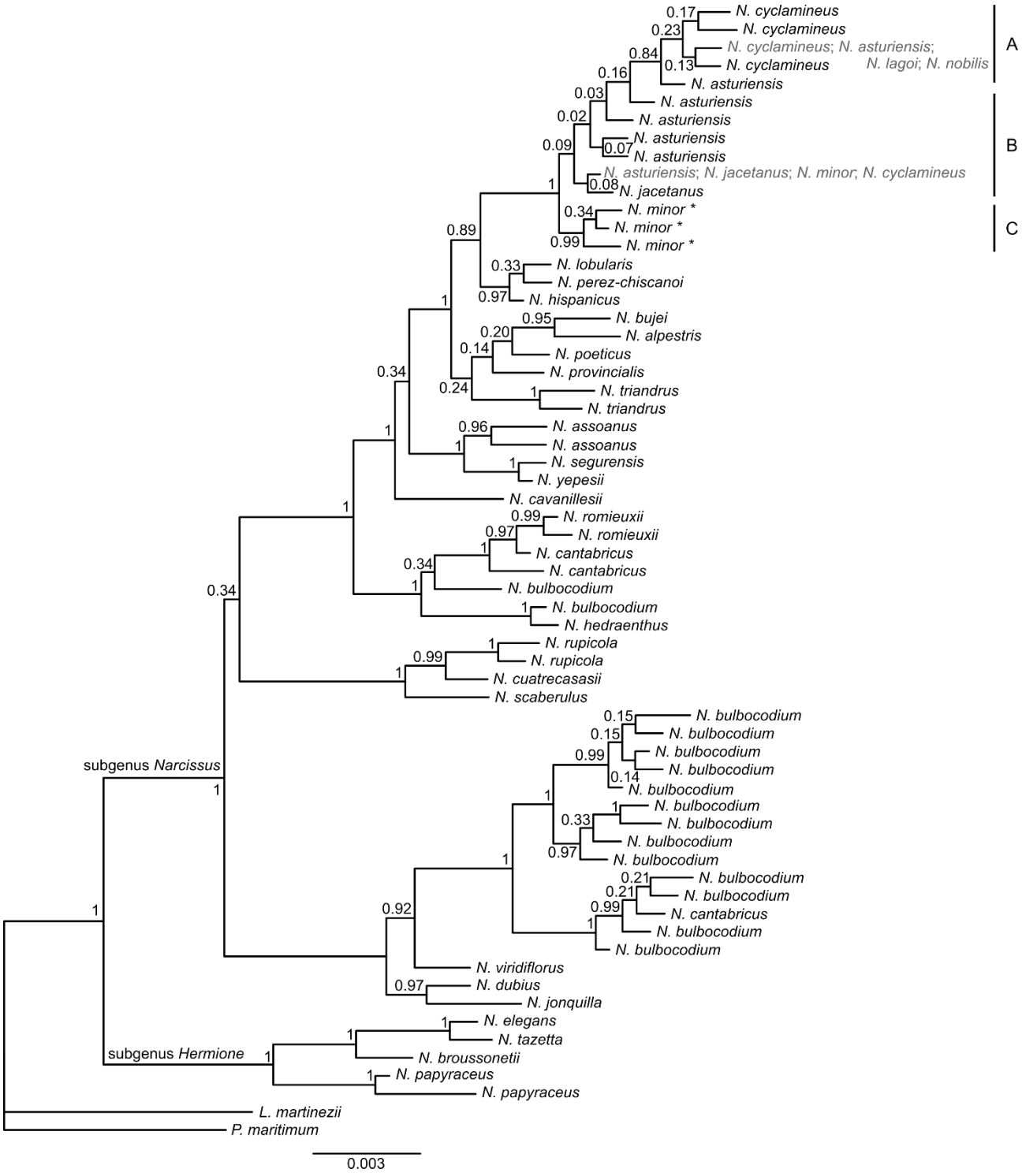


Figure 2-9 Combinable component consensus Bayesian inference tree of the *matK* and *ndhF* dataset. Posterior probabilities are shown at nodes. Scale bar shows the number of substitutions per site. Groups A, B, and C correspond to populations groupings shown in Table 2-4.

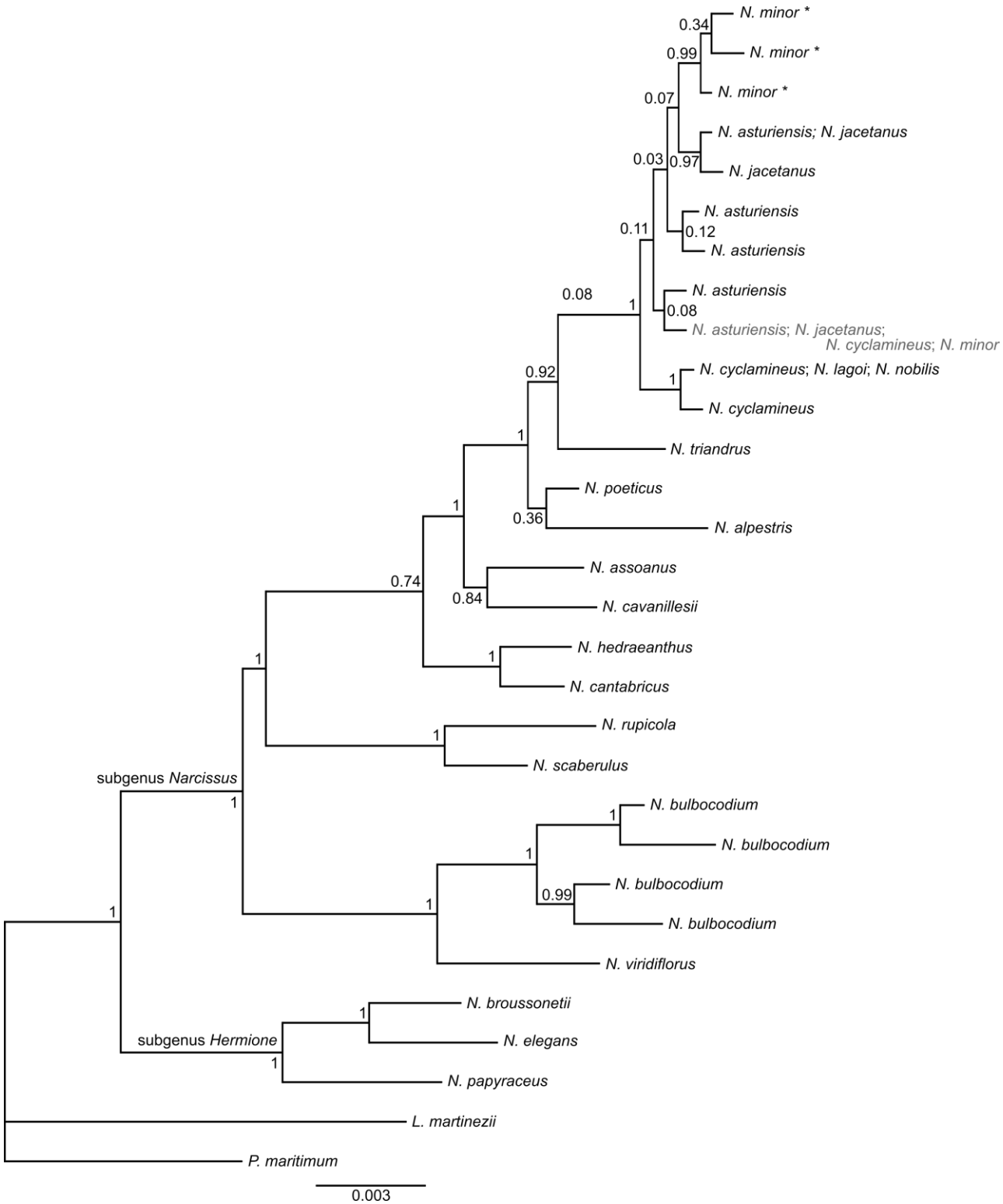


Figure 2-10 Combinable component consensus Bayesian inference tree of the *matK*, *ndhF*, and *ycf1* combined dataset. Posterior probabilities are shown at nodes. Scale bar shows the number of substitutions per site.

Table 2-4 Groups of populations with identical *matK* haplotypes, corresponding to Figure 2-9. N=Number of individuals. Population information detailed in Appendix A.

Group	N	Populations
A	167	AGA; BD18; BD19; BD26; BD35; BD54; BD55; BD56; BD57; BD60; BD61; BEM; CAN; CEB; CUE; DEZ; FRI; LER LDS; MON; PAC; PON; REG; RM; SCO; SDC; SDP; SIG; XER
B	341	ALT; Aralar; ARG; BD05-15; BD17; BD20; BD24; BD25; BD29; BD30; BD31; BD32; BD41; BDL; BRA; CDC; Entzia; GOL; ISI; LUR; PBL; PdF; PDH; PdLi; PDM; PDP; Per; PMA; POV; ROZ; SAN; SAR; SEN; SJM; SM; TEB; VDA; VM
C	35	ARA; LDS; MDS; SDP

2.4.4 Haplotype network of *matK*

The TCS network is based upon an 830bp region of the *matK* gene. The TCS identified 5 haplotypes within the study group (Figure 2-11). The majority of populations sampled correspond to H5 and H7, with 35% and 58% of the sampled populations respectively. The two haplotypes were only separated by a single base pair difference. Moreover, there was an overlap of the sampled taxa clustering in each of the two largest haplotypes. Four taxa were found clustering within H5: *N. asturiensis*, *N. cyclamineus*, *N. jacetanus*, and *N. lagoi*. Only *N. cyclamineus* corresponds to H6, found within a single population, PON. The most populous haplotype, H7, included *N. asturiensis*, *N. jacetanus, N. minor,* as well as three subsection *Pseudonarcissi* taxa not considered within the *Narcissus minor* group: *N. varduliensis*, *N. nobilis*, and *N. pseudonarcissus*. Haplotype H9 corresponds to populations collected of *N. minor* that were not flowering at the time of sampling. Clusters 1, 2, and 3 in the DAPC analysis

Figure 2-2) are shared amongst haplotypes H5-H8. Haplotype H9, corresponds to Cluster 4, identified in the DAPC analysis.

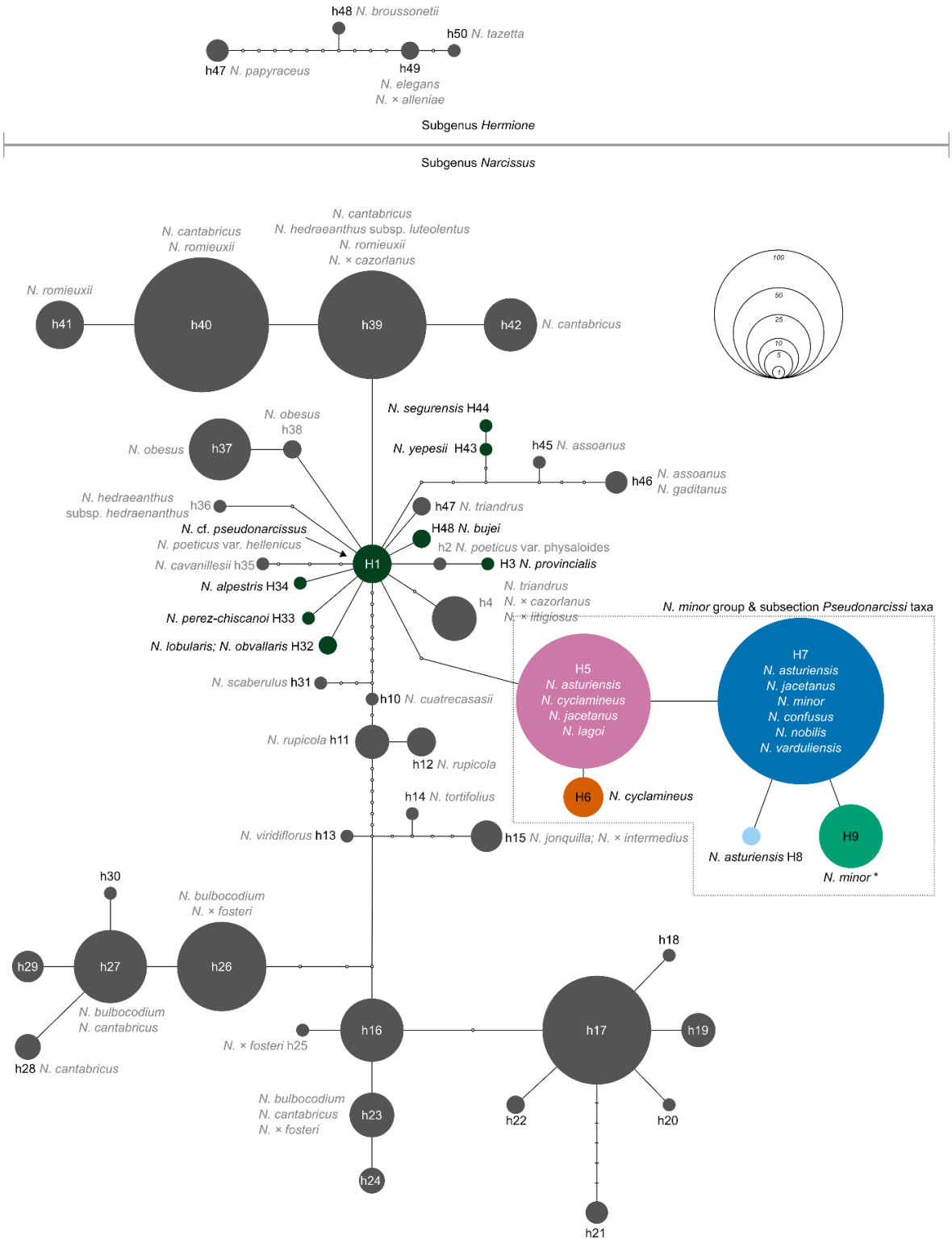


Figure 2-11 Haplotype network of *matK* sequences, haplotype shaded green contain subsection *Pseudonarcissi* taxa. Coloured haplotypes correspond to *Narcissus minor* group taxa. Haplotypes shaded grey correspond to other *Narcissus* taxa. The size of each circle is proportional to haplotype frequency. The length of the lines is not indicative of genetic distance, open circles indicate inferred haplotypes. The *matK* sequences for outgroup taxa were compiled from Könyves, David, et al. (2019).

The geographic distribution of the haplotypes within the study group is shown in Figure 2-12. Two of the populations, BRA and MDS contained more than one haplotype, whilst the remaining populations had only one haplotype. One population of *N. cyclamineus* collected in northern Portugal (PON) contained a separate haplotype from the remaining *N. cyclamineus* samples analysed here, and was the only population to carry that haplotype. The two largest haplotypes within the study group have a geographic structure. Populations belonging to haplotype H5 are mostly found west of the Galicia Leon and Cantabrian Mountains, and in northern Portugal. However, haplotype H7 was found towards the east of the distribution range, and in central Portugal. There is a contact zone within the Galicia Leon Mountains in Castile & Leon, near populations with haplotype H9. The geographic structure of *matK* haplotypes does not corroborate the geographic structure found in the DAPC analysis.

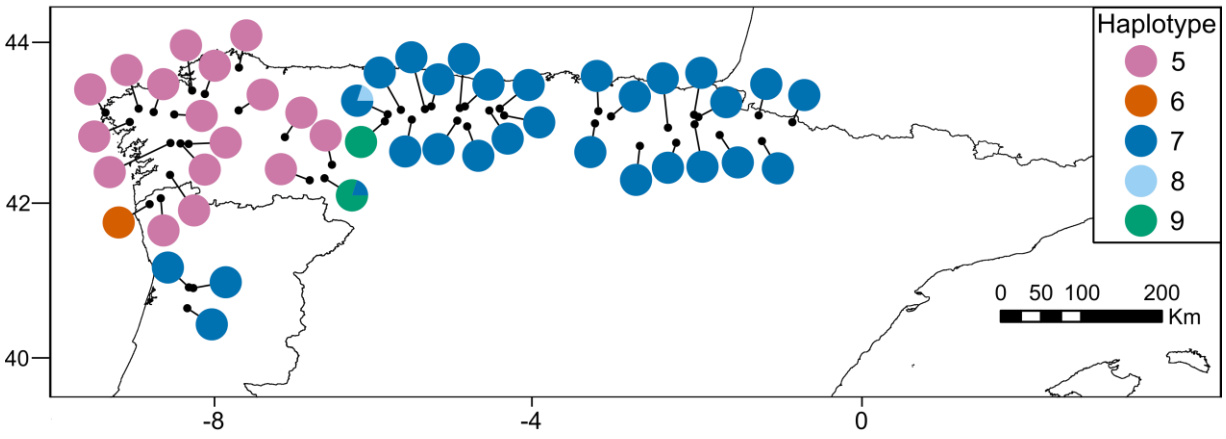


Figure 2-12 Geographic distribution of *matK* haplotypes across northern Iberian Peninsula. Colours correspond to the colours in the haplotype network (Figure 2-11).

2.5 Discussion

The phylogenetic analysis shows support for the subgeneric split of *Narcissus* into *Narcissus* and *Hermione*, in agreement with previous molecular studies (Graham and Barrett, 2004, Santos-Gally, Vargas, et al., 2012, Marques, Aguilar, et al., 2017). Several sections are not monophyletic based upon the phylogenetic analysis, including the trumpet daffodils. This supports the recent change in taxonomic rank to subsection *Pseudonarcissi* within section *Narcissus* (Marques, Aguilar, et al., 2017). The disparity between the BI trees based upon two-region and three-region datasets is likely to be due to the difference in sampling. None of the

currently recognised species in the *N. minor* group form monophyletic groups using plastid markers, corroborating previous phylogenies which have sampled taxa from the group (Graham and Barrett, 2004, Marques, Aguilar, et al., 2017, Könyves, David, et al., 2019), however this study includes sampling from the entire distribution range. Moreover, *N. cyclamineus* could not be separated from taxa within the *N. minor* group, *N. cyclamineus* is morphologically distinct, and can easily be separated from other trumpet daffodils as it has reflexed tepals (Fernandes, 1968, Webb, 1978).

The levels of genetic diversity were high across the 56 populations, indicated by the number of multilocus genotypes observed in the microsatellite data, in line with previous population level studies in *Narcissus* (Hodgins and Barrett, 2007, Simon-Porcar, Pico, et al., 2015, Könyves, David, et al., 2019). The observed heterozygotes ranged from 0.16 to 0.50, similar to the 0.09 to 0.55 found in hoop-petticoat daffodils (Könyves, David, et al., 2019) and within the range of 0.28 to 0.76 found across *Narcissus triandrus* populations (Hodgins and Barrett, 2007). The F_{st} values were low (0.29) characteristic of populations with low divergence, although lower F_{st} values (0.10) were observed amongst populations of one species, *N. triandrus* (Hodgins and Barrett, 2007).

Neither plastid DNA nor SSR data were able to separate any of the taxa described within the *N. minor* group, nor was it possible to separate *N. cyclamineus* from the taxa within the group. Morphological data have previously been used to classify this species in a monotypic sub-section *Reflexi,* within section *Pseudonarcissus* (Webb, 1978), or section *Cyclaminopsis* (Pugsley, 1933). However, previous molecular analysis has also failed to separate *N. asturiensis* and *N. cyclamineus* using plastid DNA data (Graham and Barrett, 2004, Marques, Aguilar, et al., 2017, Könyves, David, et al., 2019). Analysis based upon ITS data was able to separate *N. cyclamineus* from the *N. minor* group taxa; however, this was based upon one sample (Marques, Aguilar, et al., 2017).

Narcissus minor is recognised in morphological (Pugsley, 1933, Fernandes, 1968, Webb, 1980, Aedo, 2013) and cytological (Fernandes, 1975, Zonneveld, 2008) classifications, although the species could not be distinguished here. The only previous molecular study to include *N. minor* also failed to separate the species from other taxa in the *N. minor* group (Marques, Aguilar, et al., 2017). *Narcissus minor* has been suggested to be a possible hybrid of *N. pseudonarcissus*

and *N. asturiensis* (Blanchard, 1990, Mathew, 2002). The hybrid origin of *N. minor* could explain the molecular patterns observed in this study.

Narcissus jacetanus has been poorly defined; the species protologue separates it from *N. provincialis,* which itself is a synonym of *N. minor* (Webb, 1980, Zonneveld, 2008). *Narcissus jacetanus* was synonymised with *N. minor* in Flora Iberica, based upon morphological grounds (Aedo, 2010). It is often reported that *N. jacetanus* is limited to calcareous soils whereas *N. asturiensis* is found in acid soil however *N. asturiensis* can be found in either soil type (Fernández Casas, 1990). Given the species cannot be easily distinguished from other taxa within the *N. minor* group based upon morphological data and molecular data here, it is likely that the variation amongst populations of *N. asturiensis* has led to the recognition of *N. jacetanus*. Taxonomic inflation has also been identified within *Narcissus* section *Nevadensis,* with morphological based species not reflected in population level variation (Medrano, López-Perea & Herrera, 2014a; Jiménez *et al*, 2009).

There is clustering of both plastid sequence and microsatellite data for the sampled populations within the Castile Leon Mountains (Cluster 4 in Figure 2-2, and H9 in Figure 2-11). These populations were thought to be *N. minor* however; sampling took place whilst the plants were not in flower. To confirm identification bulbs were collected and grown within a *Narcissus* living collection at the University of Reading. The plants were small in stature with bicolored flowers, which is not found in *N. minor*, but rather appear to correspond to *N. primigenius*. *Narcissus primigenius* has been described as a smaller stature *N. nobilis* however, the species has not been fully characterised in the protologue or included in any morphological studies since publication. Further research is required to ascertain the species boundaries, as there is clear molecular evidence presented in this study to recognise *N. primigenius*, and based upon ITS, plastid and mitochondrial sequence data (Marques, Aguilar, et al., 2017). Even though there is strong molecular evidence it is likely the species has not been widely accepted due to the taxonomic inflation reported in certain groups of *Narcissus* (Jiménez *et al*, 2009; Fonseca *et al*, 2015), and the publication of superfluous names that are not data driven.

High morphological variation within populations of *N. asturiensis* has previously been recorded (Blanchard, 1990); the species is also the most widespread within the *N. minor* group. *N. asturiensis* has previously been recognised as a species on morphological grounds (Webb

1980; Fernandes 1968; Pugsley 1933). Previous molecular evidence also failed to separate *N. asturiensis* from the *N. minor* group and *N. cyclamineus* (Graham and Barrett, 2004, Marques, Aguilar, et al., 2017, Könyves, David, et al., 2019).

Despite the lack of taxonomic separation, highlighted by the placement of *N. cyclamineus* in particular, there is a clear geographic pattern in both microsatellite and plastid DNA analyses (Figure 2-6; Figure 2-12). The populations across the distribution range are clustered into two main *matK* haplotypes, one within the NW of the Iberian Peninsula and another haplotype extending both east towards the Pyrenees and south into central Portugal. Similarly, the DAPC analysis shows a cluster to the east of the distribution range towards the Pyrenees, and there is overlap of the remaining clusters for the other sampled populations. The disparity between morphology and molecular data is likely due to extensive gene flow amongst the populations. Hybridisation is prevalent and widely reported within *Narcissus* (Marques, Feliner, et al., 2010, Marques, Loureiro, et al., 2018), reflected in the polyphyletic pattern observed in some sections. The number of artificial crosses in horticulture also highlight the lack of sexual barriers amongst species of *Narcissus*.

A limit to gene flow is therefore required to explain the geographic clusters that are apparent. One example would be a difference in flowering time however, the plants within the study group overlap in flowering time as was observed during sampling. Pollinators could also provide the limit to gene flow, and have been reported to have selective pressure upon floral traits (Arroyo and Barrett, 2000). Pendulous flowers with reflexed tepals, present in *N. cyclamineus*, also feature in section *Ganymedes* (*N. triandrus*), and section *Apodanthi* (*N. albimarginatus*). However, the significance of this floral form and reason behind the occurrence in different parts of the phylogeny are currently unknown. Pollinators having a selective pressure on convergence in floral traits has been reported in section *Apodanthi* and *Ganymedes*; and the same groups of species pollinate all three species with reflexed tepals. However, further detailed pollinator studies are required to elucidate this relationship, particularly for *N. cyclamineus*. The patterns of variation observed reflect extensive pollen flow between high altitude populations and into the low-lying areas of Galicia (Figure 2-13). Pollen dispersal mediated by bees has been found to occur over 2km in fragmented landscapes (Jha and Dick, 2010) however, most of the populations sampled here are further than 2km apart,

the patterns of variation therefore reflect a more complex historical process (Lawson, van Dorp, et al., 2018). By comparison, seed movement is far more localised, the Galician-Leon Mountains appear to act as a barrier between the east and west of the distribution range.

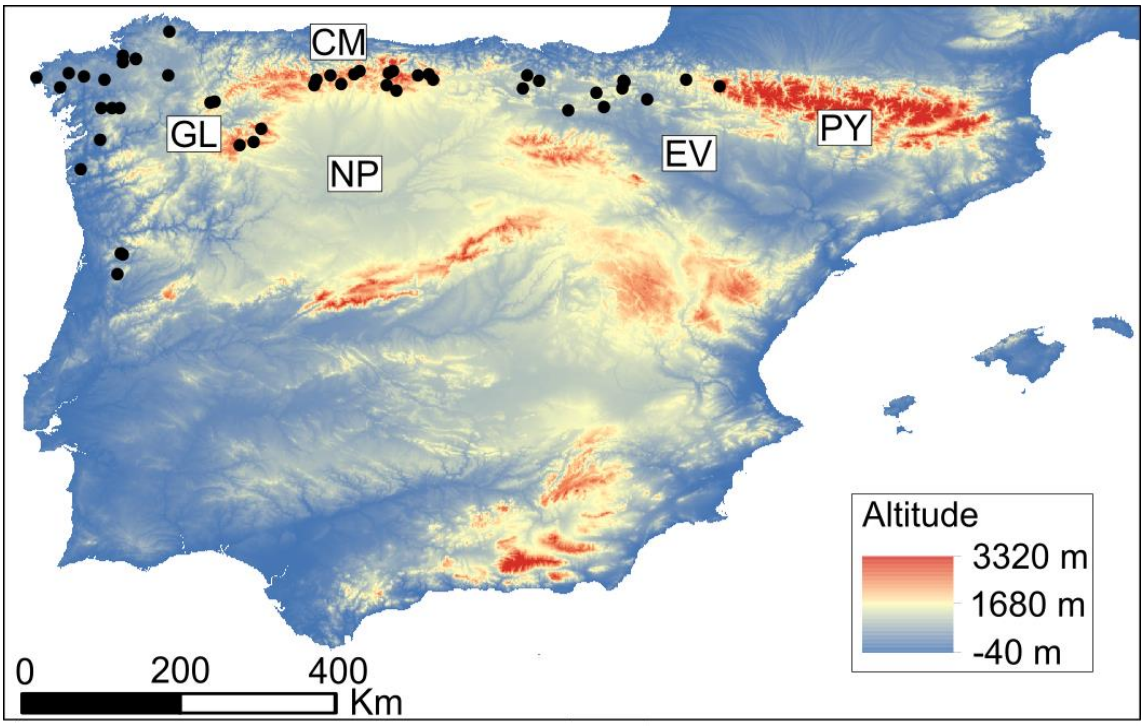


Figure 2-13 Elevation map of the Iberian Peninsula, mountain ranges are labelled as; GL = Galician-Leon Mountains ; CM = Cantabrian Mountains ; NP = North Plateau ; EV = Ebro Valley; PY = Pyrenees. Black circles represent the populations sampled.

The most likely reason to explain the clustering of panmictic populations in both DAPC and plastid DNA haplotype plots is the topography of northern Spain, notably altitude. The mountains ranges of the northern Iberian Peninsula are shown in relation to the sampled populations in Figure 2-13. For the DAPC analysis (

Figure 2-2), cluster 1 is found throughout the northern region of the Ebro Valley, that is not present any further west upon contact with the Cantabrian Mountains. Populations representing clusters 2 and 3 are found throughout the remaining distribution range whereas, cluster 4 corresponds to populations restricted to the Galician-Leon Mountains, with one exception being the population (ARA) located in the west most sampled population in the Cantabrian Mountains. Separate haplotypes were found in the contact zone of the Galician-Leon Mountains; therefore, there is strong molecular evidence to recognise *N. primigenius*.

A similar geographic structure was found amongst populations of *Daboecia cantabricus* in the same geographic region of the Iberian Peninsula (Beatty and Provan, 2013). The geographic plastid DNA structure of *D. cantabricus* populations was likely due to seed dispersal following separate glacial refugia at the Last-Glacial Maximum (LGM; c. 18-21ka) on the west coast of Galicia and within what is now the Bay of Biscay. This scenario would also explain the high similarity to the geographic structure, and patterns of variation observed in the *N. minor* group. Climatic models available for the Iberian Peninsula during the LGM found the Pyrenees and the north-west of the Peninsula to be species rich refugia for 19 tree species (Benito Garzón, Sánchez de Dios, et al., 2007), and highlight the importance of montane populations for post glaciation dispersal. Heap *et al* (2014) found niche movement in response to a changing climate occurs in an east and west direction in the Iberian Peninsula, this pattern appears to correlate with the patterns of variation shown in this study. Moreover, the phylogeographic pattern of *Meconopsis cambrica* was observed using plastid DNA and ITS sequence data (Valtueña, Preston, et al., 2012), that revealed separate clustering of populations in the Cantabrian Mountains and in the Ebro Valley. This climatic scenario that explains the patterns of variation should be explored further using climate modelling in the region to elucidate the relationship of post-glacial dispersal limited by topography as it appears to have played a role in the diversification of *Narcissus*.

The recent diversification in daffodils (Santos-Gally, Vargas, et al., 2012), with the widely reported hybridisation (Marques, Loureiro, et al., 2018) and levels of ploidy (Fernandes, 1975, Zonneveld, 2008) provide challenges in classifying the taxa within the genus. It is possible that the morphological variation, which has given rise to the wide range of taxa being described, is not reflected in the molecular variation due to the rapid diversification of the genus. Similar patterns of high morphological diversification were not reflected in molecular patterns of *Lupinus* species (Hughes and Eastwood, 2006).

Microsatellites and cpDNA sequencing identified a geographic separation in the *Lotus corniculatus* complex, between European and Asian distributed individuals, with species boundaries within the two geographic clusters unclear due to hybridisation (Kramina, Meschersky, et al., 2018). Consequently, Kramina et al. (2018) used a combination of morphological and molecular evidence to separate species in *Lotus.* Within *Narcissus*

papyraceus high morphological variation of floral and vegetative characters did not correspond with the molecular population variation of the species (Pérez-Barrales, Pino, et al., 2009). Given that classification and identification makes use of morphological data, it would be worthwhile to identify the how certain morphologies such as reflexed tepals have originated in separate parts of the phylogeny, and how they persist despite a lack of molecular separation. It would therefore be ideal to sequence transcriptomes of representatives of *Narcissus*.

The taxonomic conclusions from this study reveal that none of the taxa within the *N. minor* group can be separated on the molecular evidence gathered. Furthermore, the members of the *N. minor* group cannot be separated from *N. cyclamineus,* although *N. primigenius* is supported by molecular data. Given *N. cyclamineus* is an IUCN red listed species with habitat threats, it is paramount to ascertain the species boundaries through further study. Type specimens of both *N. cyclamineus* and *N. asturiensis* were only recently selected whilst producing the *Narcissus* account in Flora Iberica (Aedo, 2010), highlighting the lack of taxonomic research into the group. With the publication of a *N. poeticus* plastid genome (Könyves, Bilsborrow, et al., 2018) highly variable regions within the chloroplast genome, high-copy nuclear and mitochondrial DNA regions can be identified for further study. It is essential to use multiple plant genomes to unpick the patterns of prevalent hybridisation within *Narcissus,* and therefore harnessing bigger data sets that Next-Generation sequencing provide. This could be key to elucidate the molecular variation for circumscription of species boundaries.

Chapter 3 Reference chloroplast genome of *Narcissus poeticus*

Published – Mitochondrial DNA Part B

The complete chloroplast genome of *Narcissus poeticus* L. (Amaryllidaceae: Amaryllidoideae)

Kálmán Könyves[a*], Jordan Bilsborrow[b], John David[a] and Alastair Culham[b]

[a]Royal Horticultural Society Garden Wisley, Woking, United Kingdom; [b]Herbarium, School of Biological Sciences, University of Reading, Reading, United Kingdom;

Kálmán Könyves; kalmankonyves@rhs.org.uk; Royal Horticultural Society Garden Wisley, Woking, Surrey GU23 6QB, United Kingdom

Word count: 480

The complete chloroplast genome of *Narcissus poeticus* L. (Amaryllidaceae: Amaryllidoideae)

Abstract

The first complete chloroplast genome sequence for *Narcissus* is assembled and annotated in this study. The total length of the *N. poeticus* chloroplast genome is 160,099 bp and comprises the large single copy (LSC) spanning 86,445 bp, the small single copy (SSC) spanning 16,434 bp, and two inverted repeat regions each of 28,610 bp length. The truncated copy of *ycf1* before the junction between IR_B and SSC was 1,277 – 2,428 bp longer than in other included Asparagales samples. A potential pseudogene, *cemA*, was also identified. This is the first reported plastome for Amaryllidaceae subfamily Amaryllidoideae.

Keywords: *Narcissus*; Amaryllidaceae; plastome; *ycf1; cemA*

Narcissus L. is a member of Amaryllidaceae, a family comprising several horticulturally important plant genera . *Narcissus poeticus* is the type species of the genus, however its placement among other daffodils remains unresolved (Marques et al. 2017). Currently there are no published daffodil chloroplast genomes and only few genomes are available for Amaryllidaceae, none for subfamily Amaryllidoideae. Here we report the complete chloroplast genome sequence of *N. poeticus* to address this and to identify new regions of genomic variability.

Leaf material was collected in silica gel from *N. poeticus* grown at RHS Garden Wisley, UK (51.312695° N, 0.476724° W). Herbarium voucher specimen was deposited at WSY (WSY0108940). Total genomic DNA was extracted using the QIAGEN DNeasy Plant Mini Kit (QIAGEN, Manchester, UK). Library development and 150bp PE sequencing on 1/16 of an Illumina HiSeq 4000 lane was done at the Oxford Genomics Centre (Oxford, UK). The chloroplast genome was assembled with Fast-Plast v1.2.6 (McKain and Wilson 2017) and NovoPlasty v2.7.0 (Dierckxsens et al. 2016). Fast-Plast assemblies were run with all 38M reads. Reads were trimmed to remove NEB-PE adapter sequences. Bowtie reference index was built with Asparagales chloroplast genomes included in Fast-Plast. For the NovoPlasty assembly, adapters were trimmed with Trimmomatic v0.36 (Bolger et al. 2014) using the same adapter sequences. A *ndhF* sequence of *N. poeticus* (KT124416) was used as the starting seed and

memory was limited to 8 Gb. All other parameters were unchanged. Junctions of the inverted repeats were confirmed by Sanger sequencing.

The complete chloroplast genome was annotated using Verdant (McKain et al. 2017) and corrected by comparing it with published annotations (*Hesperoyucca whipplei* – KX931459; *Hosta ventricosa* – KX931460; *Yucca schridigera* – KX931469; *Cocos nucifera* – KF285453). The *N. poeticus* cpDNA genome sequence was aligned to 13 published Asparagales plastome sequences using MAFFT (Katoh and Standley 2013). A maximum likelihood estimate was conducted with RAxML v8.2.11 (Stamatakis 2014) within Geneious v11.1.5 (http://www.geneious.com, Kearse et al. 2012) using model GTR+G and 1000 bootstrap replicates.

The chloroplast genome sequence of *N. poeticus* (MH706763) is 160,099 bp, comprises the large single copy (LSC) spanning 86,445 bp, the small single copy (SSC) spanning 16,434 bp, and two inverted repeat regions each of 28,610 bp length. The junction between IR_B and SSC (J_{SB}) is 34 bp within the *ndhF* gene. The junction between SSC and IR_A (J_{SA}) is within the *ycf1* gene, which is 5,346 bp long, of which 2,737 bp lies in the inverted repeat. Therefore, there is a 2,737 bp long truncated version of *ycf1* at J_{SB}. This is 1,277 – 2,428 bp longer than the truncated copy of *ycf1* in other Asparagales sequences analyzed here, identified by IRScope (Amiryousefi et al. 2018). A poly-A region, at the start of *cemA* shifts the ORF potentially inactivating this gene. The Amaryllidaceae samples form a clade with *N. poeticus* sister to *Allium cepa* (Figure 1) topologically consistent with Seberg et al. (2012).

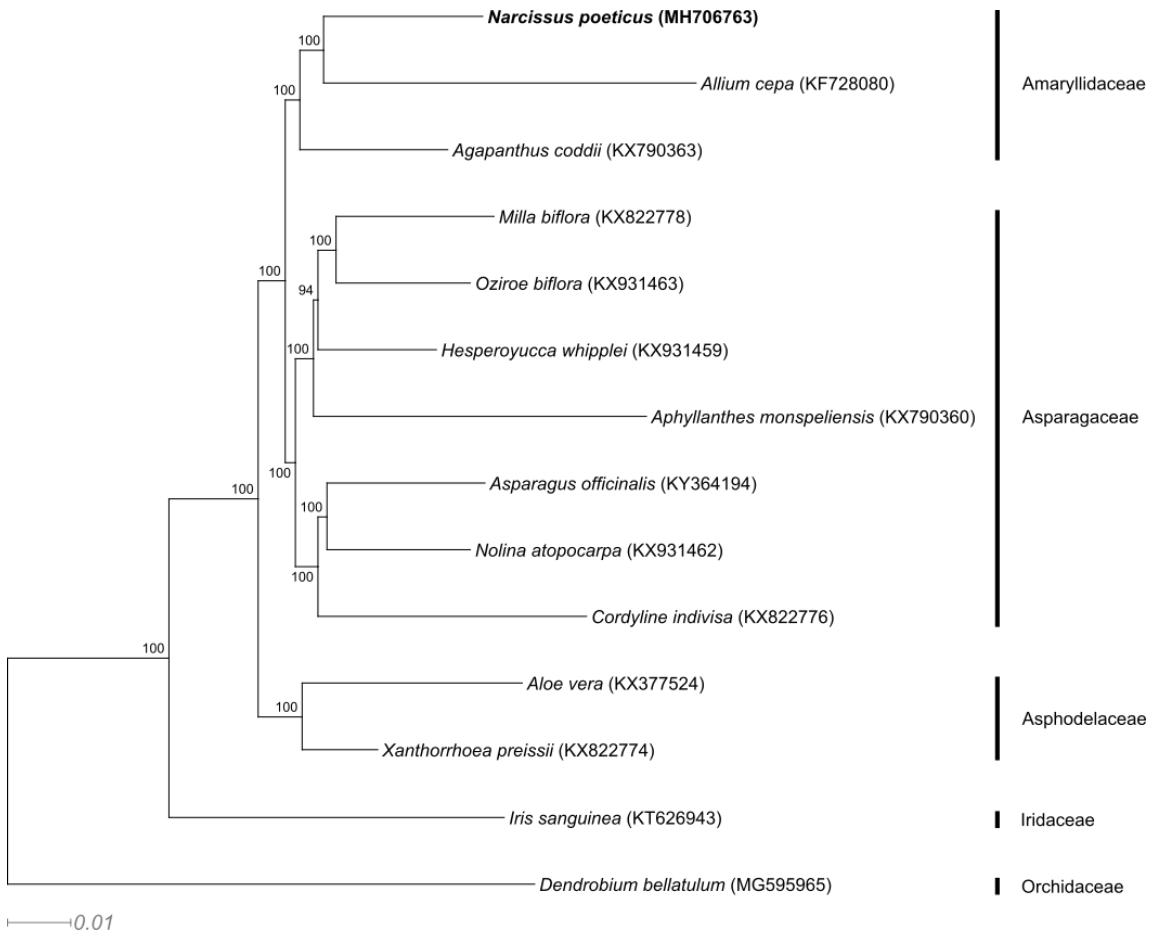


Figure 3-1 RAxML output tree with bootstrap consensus values based on 14 complete chloroplast genome sequences. The numbers at each node indicate bootstrap support. GenBank accession numbers are given in brackets. Text in bold shows the chloroplast genome developed in this study. Families of the sampled taxa are shown on the right.

Acknowledgments

We thank the High-Throughput Genomics Group at the Wellcome Trust Centre for Human Genetics (funded by Wellcome Trust grant reference 090532/Z/09/Z) for the generation of sequencing data.

Disclosure statement

The authors report no conflict of interest and are alone responsible for the content and writing of the paper.

ORCID

Kálmán Könyves https://orcid.org/0000-0002-8303-1544

Jordan Bilsborrow https://orcid.org/0000-0002-4404-9514

John David https://orcid.org/0000-0003-0557-9886

Alastair Culham https://orcid.org/0000-0002-7440-0133

Chapter 4 Comparative plastid genomics in *Narcissus*

4.1 Introduction

The genus *Narcissus* L., in the family Amaryllidaceae, includes approximately 30 species, although the number of accepted taxa ranges from 16 (Baker, 1875) to 160 (Haworth, 1831). The genus is found throughout Europe and northern Africa, although the centre of diversity for the genus is the Iberian Peninsula (Blanchard, 1990). Daffodils have been cultivated and traded as ornamentals for centuries (Rivera Nuñez, Obón De Castro, et al., 2003), with intensive breeding programmes over the past 100 years, in particular, giving rise to more than 30,000 registered daffodil cultivar names (Kington, 2008). In order to understand the variation in cultivation it is key to underpin the taxonomy of the wild species, as they will encompass the variation of daffodil cultivars.

Daffodils have a clear horticultural importance; despite this, the taxonomy is in flux. Most recent classifications have been based upon regional treatments (Webb, 1980, Aedo, 2013, Fennane, Ibn Tattou, et al., 2014). Generally, the genus is split into subgenera *Narcissus* and *Hermione* and further classified into 10-14 sections (Marques, Aguilar, et al., 2017). Based upon morphological evidence, the genus includes 10 sections (Fernandes, 1968, Webb, 1980, Blanchard, 1990), although cytological evidence provides evidence for 12 sections (Zonneveld, 2008). Splitting section *Jonquillae* and *Pseudonarcissi* based upon cytological evidence was supported using molecular data, and a further three sections proposed (Marques, Aguilar, et al., 2017). Furthermore, a change in taxonomic rank of section *Pseudonarcissi* was proposed to account for the polyphyly of the section, to incorporate the placement of section *Narcissus*. The incongruence between sections based upon morphological, cytological, and molecular data is shown in Table 4-1. Moreover, there is disagreement about the taxonomic rank at which many taxa within genus should be classified, due to wide morphological variation within species.

Table 4-1 Comparison of *Narcissus* sectional classification, comparing morphological, cytological, and phylogenetic studies.

Morphological section classification (Fernandes, 1968)	**Somatic DNA content section classification (Zonneveld, 2008)**	**Phylogenetic section classification (Marques, Aguilar, et al., 2017)**
Braxireon	*Braxireon*	*Braxireon*
Serotini	*Serotini*	*Serotini*
Aurelia	*Tazettae*	*Aurelia*
Tazettae	*Tazettae*	*Tazettae, Dubii, Angustifolii*
Narcissus	*Narcissus*	*Narcissus* subsect. *Narcissus*
Jonquillae	*Jonquillae, Juncifolii*	*Jonquillae*, *Juncifolii*
Apodanthi	*Apodanthi*	*Apodanthi*
Ganymedes	*Ganymedes*	*Ganymedes*
Bulbocodii	*Bulbocodii*	*Bulbocodii*, *Meridionalis*
Pseudonarcissi	*Pseudonarcissi, Nevadensis*	*Narcissus* subsect. *Pseudonarcissi*, *Nevadensis*

To date the most comprehensive *Narcissus* phylogeny (Marques, Aguilar, et al., 2017), is based upon 4,167 bp of DNA sequence data per sample, from three of the plant genomes (1285 bp cpDNA, 2000 bp mtDNA, and 882 bp nuclear ribosomal DNA). The remaining available *Narcissus* phylogenies are based only upon a combination of four plastid regions, *matK*, *ndhF*, and *trnL-F*, and *trnT-F* (Graham and Barrett, 2004, Santos-Gally, Vargas, et al., 2012, Könyves, David, et al., 2019). With the rapid uptake in Next Generation sequencing methods (NGS), it is now possible to acquire much larger datasets for phylogenetic study. The assembly of plastid genomes is becoming increasingly popular for many plant groups, and is no longer limited to model organisms (Straub, Parks, et al., 2012). Given the relatively small amount of data previously collected for molecular studies of *Narcissus*, genome skimming can be used to increase data quantity. Moreover, genome skimming can be used to identify variable regions that can be used to elucidate species complexes in the genus, such as the *N. minor* group.

Genome skimming involves shallow sequencing of genomic DNA. Despite the shallow sequencing, high copy regions of the genome, such as; plastid, mitochondria, and nuclear ribosomal DNA can have a read depth of hundreds of copies (Straub, Parks, et al., 2012). This provides the access to larger datasets than those used in traditional phylogenetic studies (Dodsworth, 2015). For non-model organisms, or those with complex genomes, assembly of whole genomes remains problematic however, NGS can be used to evaluate sequence variation especially of organelles, without the requirement for baseline knowledge of the genome being studied (Cronn, Knaus, et al., 2012). This means Genome skimming can help overcome the challenges with species identification studies, by identifying variable regions of the plant genome. The suggested DNA barcode regions, *matK* and *rbcL* (CBOL Plant Working Group, 2009) have limitations with difficulties in cross-amplification and low divergence for species level study respectively (Li, Yang, et al., 2015). With the difficulty in finding universal DNA barcodes for plants, genome skimming has been suggested as the mechanism to gain 'extended barcodes' (Coissac, Hollingsworth, et al., 2016). This approach has been applied to a large-scale barcoding project PhyloAlps, to sequence 4,000 different species from the European Alps flora (Coissac, Boyer, et al., 2018).

Plastid genomes have also been used to elucidate boundaries between micro-species (Salih, Majeský, et al., 2017), demonstrating the utility of plastid genomics for discriminating between closely related taxa, in species groups where hybridisation is not prevalent. However, by sequencing high copy regions of total genomic DNA the raw data are also mineable for further nuclear ribosomal and mitochondrial DNA regions available in high copy number (Ripma, Simpson, et al., 2014). Therefore, as plastomes reflect maternal lines of inheritance, the biparentally inherited nuclear genome regions should be used to resolve challenges from hybridisation, introgression and incomplete lineage sorting (Vargas, Ortiz, et al., 2017).

Whole plastid genomes were recently used to build a phylogeny with representatives from all 77 monocotyledon families, using 77 coding plastid genes to ascertain the evolutionary position of each family (Givnish, Zuluaga, et al., 2018). Givnish et al. (2018) found both phylogenetic support and branch ascertainment increased with number of genes sampled across the plastid genome, highlighting the utility of plastid genomes for phylogenetic study. High through-put sequencing has revolutionised the ease of access to genomic data,

facilitating the move away from analysis of a handful of genes, to much larger genomic datasets (Twyford and Ness, 2017).

Similarly, large-scale sequencing projects have used Next-Generation sequencing to provide large quantities of data for projects involving phylogenetic diversity. To date 1,341 transcriptomes have been published for the 1kp transcriptome project (Matasci, Hung, et al., 2014), which has subsequently been superseded by the 10kp project (Cheng, Melkonian, et al., 2018). The 10kp project builds upon 1kp with the aim to assemble whole genome datasets for samples from every major clade in the plant kingdom by 2022. The genome size of *Narcissus* (2C values 13.7-67.7) and genome complexity with ploidy variation (Fernandes, 1975, Zonneveld, 2008), means contributing to such large-scale projects is problematic, as sequence assembly and data analysis of complex genomes is challenging. However, a pipeline was recently developed to construct the first *Tulipa* genome (2C values 32.1-120.9) from long sequencing reads (Oxford Nanopore Technologies, 2018), although this was a diploid taxon removing some of the complexity, and still remains expensive per sample.

In this study, we present a further nine plastid genomes of *Narcissus* representing eight sections of the genus, and two outgroup taxa *Pancratium maritimum* L. and *Lapiedra martinezii* Lag. Here we provide a comparative analysis of the sequenced plastid genomes, identifying highly variable regions within the genome and simple sequence repeats (SSRs), which provide a bank of data for a future taxonomic revision. Furthermore, we provide initial phylogenetic analysis to test sectional placement within *Narcissus*, using plastid DNA regions and mined nuclear DNA.

4.2 Objectives

The aim of this chapter is to assemble nine further plastid genomes for *Narcissus*, and two outgroup species to provide a comparative study. This allows variable regions of the plastid genome to be identified for future study to elucidate species boundaries in *Narcissus*. Thus, the objectives are to;

1. To assemble a further nine plastid genomes for *Narcissus* and undertake a comparative study with two outgroup species.

2. To produce an initial whole plastid genome phylogeny for *Narcissus*, a phylogeny based upon mined nuclear ribosomal sequence data and a combined phylogeny of plastid and nuclear ribosomal DNA.
3. To identify variable regions of the plastid genome both in gene sequences and microsatellites, for future use.

4.3 Materials and methods

4.3.1 Sample collection and DNA extraction

Nine leaf samples were collected from the *Narcissus* living collection at the University of Reading (Table 4-2), the sampling was chosen to represent the broadest possible coverage of the phylogenetic tree. Leaf samples were collected and frozen in liquid nitrogen and subsequently stored at -80°C until DNA extraction. Two outgroup species: *P. maritimum* and *L. martinezii*. were collected from RHS Garden Wisley. Voucher specimens for each of the species sampled were deposited at **RNG** and **WSY**.

Table 4-2 *Narcissus* sections sampled and the species chosen to represent them.

Narcissus **section/subsection**	**Species**
Apodanthi	*N. rupicola*
Bulbocodii	*N. bulbocodium*
Meridionalis	*N. cantabricus*
Ganymedes	*N. triandrus*
Jonquillae	*N. viridiflorus*
Juncifolii	*N. assoanus*
Narcissus	*N. poeticus*
Pseudonarcissi	*N.* cf. *jacetanus*
Pseudonarcissi	*N.* cf. *nobilis*
Tazettae	*N. papyraceus*

Total genomic DNA was extracted using a QIAGEN DNeasy Plant Mini Kit (QIAGEN, Manchester, UK). Quality and quantification of extracted DNA was observed using 0.8% agarose gel electrophoresis, with Hyperladder™ 1kb (Bioline, London, UK) for reference. Quantification of DNA was also examined using spectrometry (NanoDrop Lite; Thermo Fisher Scientific, Paisley, UK), and fluorescence spectroscopy (Qubit 4 Fluorometer; Thermo Fisher Scientific).

4.3.2 Genome sequencing, assembly, annotation, and analysis

Genomic DNA was submitted to the Oxford Genomics Centre (Oxford, UK) for library preparation and sequencing. Libraries were prepared followed by 150bp paired-end sequencing on 1/16 of an Illumina HiSeq 4000 lane, per sample.

Raw sequencing reads were used to construct plastid genomes in Fast-Plast v1.2.6 (McKain and Wilson 2017) and NovoPlasty v2.7.0 (Dierckxsens, Mardulyn, et al., 2016). Fast-Plast assemblies were run with 5, 10, 20 million reads and all sequenced reads. Reads were trimmed to remove NEB-PE adapter sequences. A Bowtie reference index was built using Asparagales plastid genomes included within Fast-Plast. Adapters were trimmed using Trimmomatic v0.36 for the NovoPlasty assembly, using the same adapter sequences. An *ndhF* sequence of *N. poeticus* (KT124416) was used as the starting seed. To check consensus each of the plastid genome assemblies per sample was aligned using MAFFT v7.388 with default settings (Katoh and Standley, 2013) in Geneious v11.1.5 (Kearse, Moir, et al., 2012). Additionally, primers were designed to sequence through the LSC-IRb, LSC-IRa, SSC-IRa, SSC-IRb junctions to confirm the correct assembly of genomes. Sanger sequencing of partial genes was completed for the following: *matK*, *ndhF*, *ycf1*, *atpB-rbcL*, *atpH-F*, *psbI-K*, *rpl16*, and *rpoB*. Sequencing gene regions were subsequently mapped to each of the genomes to confirm assembly.

All sequenced genomes were annotated and checked using the published *N. poeticus* plastid genome (Könyves, Bilsborrow, et al., 2018). The investigation of start and stop codons was completed manually to ensure the annotations were in-frame for the remaining plastid assemblies detailed in this chapter.

Nuclear DNA gene assembly was completed as in Ripma et al. (2014b). Raw sequencing reads were mapped to a *Narcissus poeticus* partial ITS2 gene sequence (KY992421) for each sample

in Geneious. The Geneious mapping protocol was used and set to 100 iterations to build scaffolds from the ITS2 base sequence. Consensus sequences were then generated from the mapped sequence reads. The nuclear DNA regions were aligned using MAFFT with default settings in Geneious.

4.3.3 Phylogeny Inference

For the phylogenetic analysis, both whole plastid genomes were used and coding regions extracted. The 77 coding regions were concatenated for each sample. A further dataset was concatenated to include the coding regions of each plastid genome and the nuclear DNA consensus sequence. Each partitioned plastome genome dataset was aligned using MAFFT in Geneious. Maximum Likelihood analysis was implemented using the RAxML plugin (Stamatakis, 2014) in Geneious. Bootstrap analysis was completed of 1000 replicated was undertaken and search for the best-scoring ML tree.

4.3.4 Simple sequence repeats

Simple sequence repeats (SSRs) were identified using the MISA Perl script (Beier, Thiel, et al., 2017). The minimum thresholds within the configuration file were set to ten repeat units for mononucleotide SSRs, five repeat units for dinucleotide SSRs, four repeat units for trinucleotides, and three repeat units for tetra-, penta-, and hexanucleotide SSRs. The location of the identified SSRs was identified in Geneious.

4.4 Results

4.4.1 Genome sequencing and assembly

In the nine *Narcissus* species sampled in this study, 34,383,254 to 51,526,524 paired end raw sequencing reads were generated with an average read length of 150 bp. All the assemblies were successful when a 5 million read subsample of the raw data was used. Sanger sequencing confirmed the sequence of the plastome junctions and a further eight partial gene regions.

The plastid genome size for the type species of the genus, *Narcissus poeticus*, was 160,099 bp in length (Figure 4-1). The assembled genomes ranged from 157,460 bp in *N. viridiflorus* to 160,347 bp in *N. rupicola* (Table 4-3). Each region within the quadripartite structure varied in size, including LSC region ranging from 85,278-86,525 bp, the SSC region ranging from 16,391-

18,229 bp, and the IR region ranging from 26,746-28,643 bp. The overall GC content of the whole genomes was between 37.8%-38%.

The outgroup species, *P. maritimum* and *L. martinezii,* have a plastid genome size of 160,123 bp and 159,022 bp respectively. The length of the IR and SSC regions were similar in length for *N. viridiflorus* and *L. martinezii* (Table 4-3). A total of 138 coding genes were identically annotated in the same order, consisting of 77 protein coding genes, 38 tRNA genes, and 8 rRNA genes.

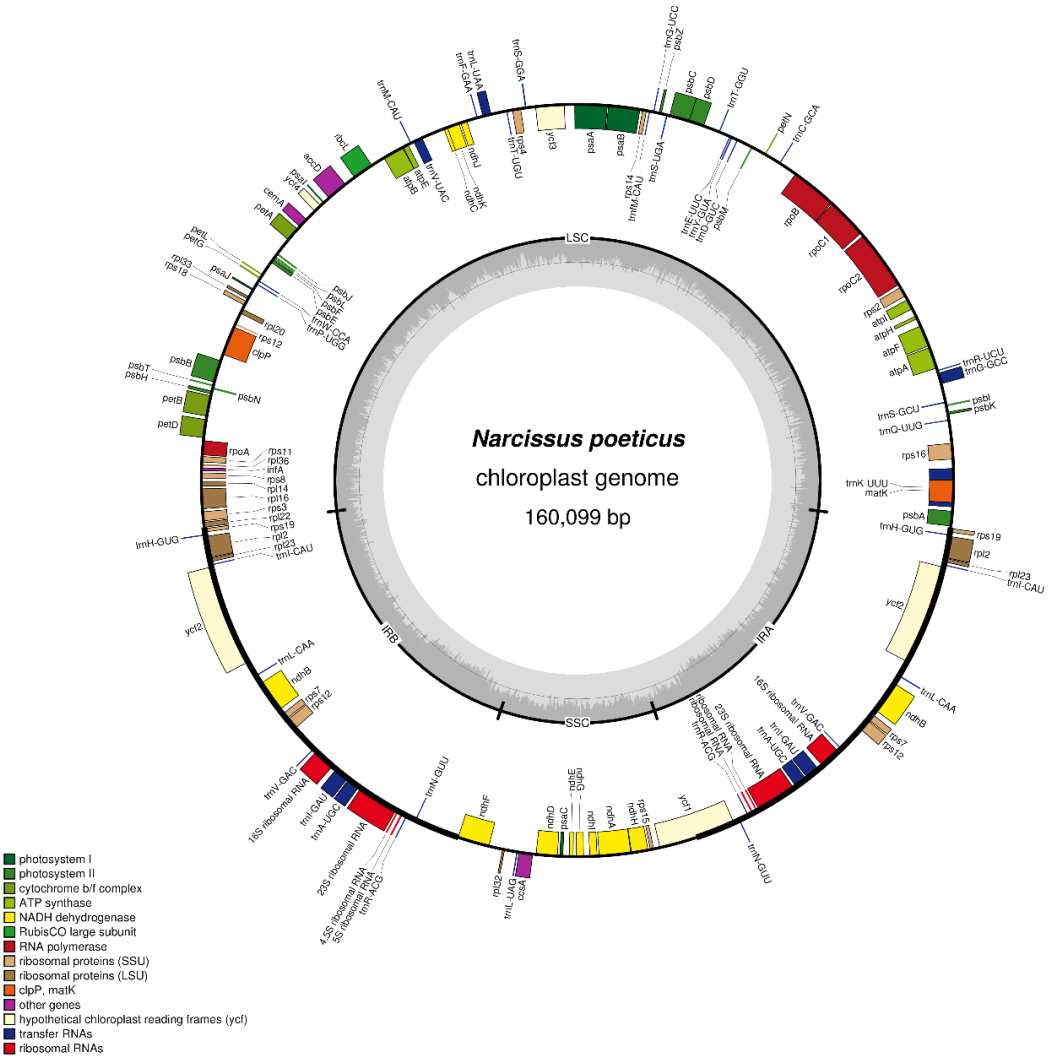


Figure 4-1 Gene map of *Narcissus poeticus* plastid genome. Genes shown on the outside are transcribed counter-clockwise, and genes inside are transcribed clockwise. Genes belonging to different functional groups are colour coded. The grey circle (inset) shows GC content.

Table 4-3 Basic characteristics of the 12 plastid genomes included within this study.

Species	**Total cpDNA size (bp)**	**Length of LSC (bp)**	**Length of IR (bp)**	**Length of SSC (bp)**	**Total no. of genes**	**Total no. protein coding genes**	**Total no. of tRNA**	**Total no. of rRNA**	**GC content**			
									Overall (%)	**LSC (%)**	**IR (%)**	**SSC (%)**
Narcissus assoanus	159,585	86,080	28,557	16,391	134	77	38	8	37.9	36.0	42.5	32.1
Narcissus bulbocodium	157,971	85,365	28,057	16,492	134	77	38	8	37.9	36.1	42.4	32.0
Narcissus cantabricus	158,172	85,278	28,245	16,404	134	77	38	8	37.9	36.0	42.5	31.9
Narcissus jacetanus	159,716	86,040	28,544	16,588	134	77	38	8	37.8	35.9	42.5	31.9
Narcissus nobilis	159,735	86,021	28,544	16,626	134	77	38	8	37.8	36.0	42.5	31.9
Narcissus papyraceus	159,781	86,396	28,481	16,423	134	77	38	8	38.0	36.1	42.5	32.4
Narcissus poeticus	160,099	86,445	28,610	16,434	134	77	38	8	37.8	35.9	42.4	31.9
Narcissus rupicola	160,347	86,453	28,610	16,674	134	77	38	8	37.9	35.9	42.5	32.1
Narcissus triandrus	160,219	86,525	28,643	16,408	134	77	38	8	37.8	35.8	42.4	32.0
Narcissus viridiflorus	157,460	85,739	26,746	18,229	134	77	38	8	37.9	36.0	42.2	32.0
Lapiedra martinezii	159,022	86,756	26,507	18,540	134	77	38	8	37.7	35.8	43.0	31.9
Pancratium maritimum	160,123	86,393	28,507	16,716	134	77	38	8	37.8	45.9	42.4	32.0

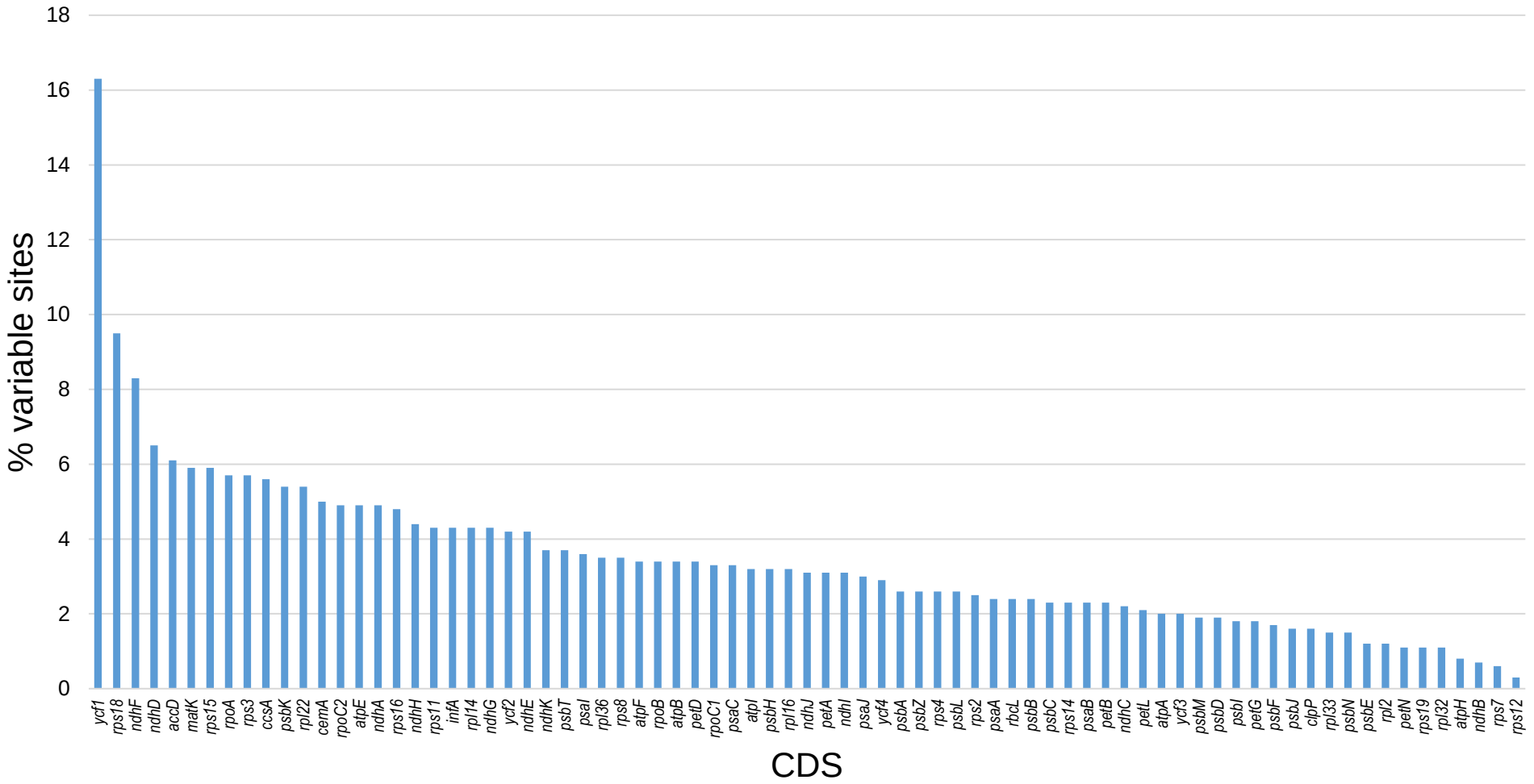


Figure 4-2 Bar plot of percent variable sites for the 77 coding genes across the 10 *Narcissus* plastid genomes.

Of all the coding genes recovered from the annotations, *ycf1* was found to be the most variable with 16.3% variable sites across the genus (Figure 4-2). Overall, there was a range from 0.3% to 16.3% variable sites for all protein coding genes. The next most variable region was *rps18*, 12 regions in total had over 5% variable sites across the plastid genome.

In five of the *Narcissus* species and *Pancratium*, *ndhF* crosses the IRb-SSC junction, by: 7 bp in *N. bulbocodium* and *N. rupicola,* 30bp in *N. viridiflorus,* 31bp in *N. triandrus,* and 34 bp in *N. poeticus* (Figure 4-3). The length variation of *ndhF* is between 2,193 bp in *N. viridiflorus* and 2,260 bp in *N. triandrus*.

The border of the SSC-IRa junction is crossed by *ycf1*, with length variation between 5,535 bp in *N. rupicola* and 5,301 bp in *N. assoanus* (Figure 4-3). For all samples within this study there is a pseudo-*ycf1* gene at the junction J_{SB} ranging from 922-2795 bp. The J_{SB} junction for *N. viridiflorus* is crossed by a 922bp truncated version of *ycf1,* which is similar to the 991bp truncated version found in *Lapiedra*. The remaining samples contained a truncated version of *ycf1* of more than 2,641 bp.

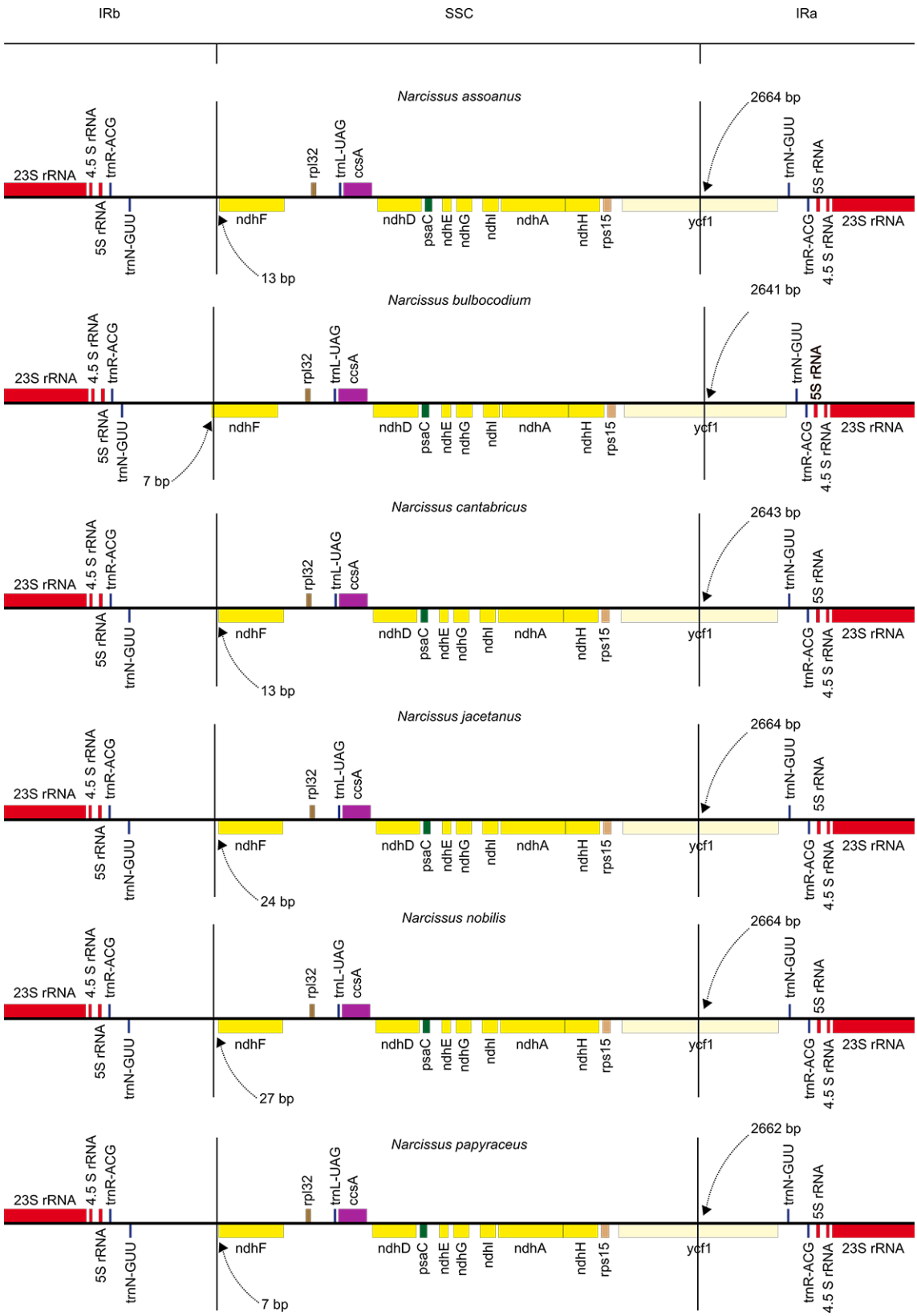


Figure 4-3 Comparison of IRb, SSC, and IRa region borders among 10 *Narcissus* species, and 2 outgroup taxa; *Pancratium maritimum*, and *Lapiedra martinezii*.

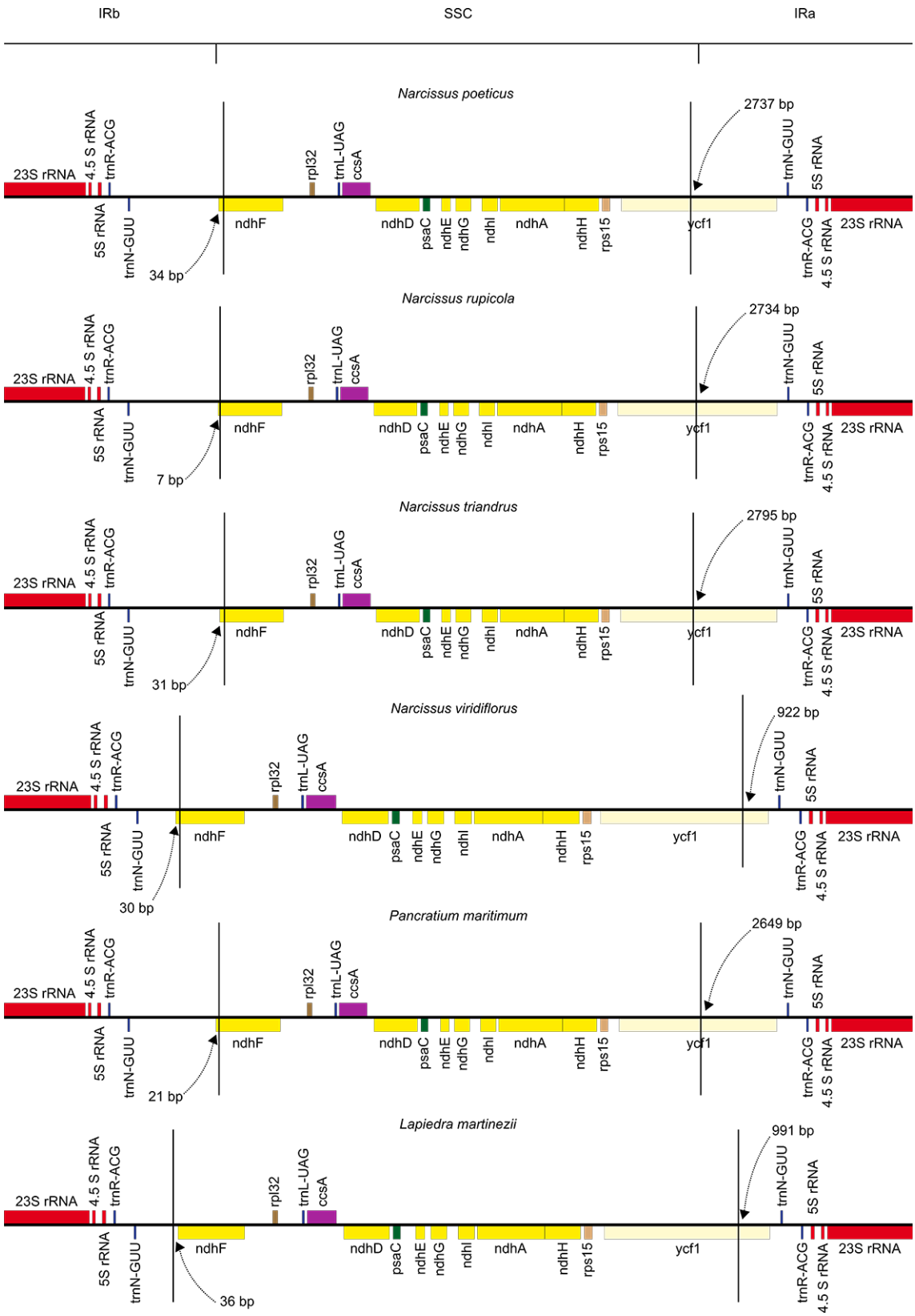


Figure 4-3 (cont.) Comparison of IRb, SSC, and IRa region borders among 10 *Narcissus* species, and 2 outgroup taxa; *Pancratium maritimum*, and *Lapiedra martinezii*.

4.4.2 Phylogenetic trees

The aligned data set for the whole plastid genomes consists of 164,211 bp, compared with 79,999 bp in the data set that only contained concatenated coding sequences. Both phylogenies built using plastid DNA data (Figure 4-4) are topologically consistent. There is support for a subgenera split to *Narcissus* and *Hermione*. The taxa included to represent subsection *Pseudonarcissi* form a monophyletic clade, sister to section *Ganymedes*. Section *Bulbocodii* is found within a clade including section *Jonquillae* that is sister to section *Apodanthi*.

The nuclear ribosomal scaffolds constructed from a partial ITS2 gene provided an alignment of 6,876 bp. The phylogeny based upon nuclear ribosomal DNA (Figure 4-5) corroborates the subgenus division of *Narcissus*, however the placement of some sections is different when compared to the phylogeny based on plastid DNA. Subsection *Pseudonarcissi* here is sister to subsection *Narcissus*, section *Bulbocodii* is sister to section *Meridionalis* rather than section *Jonquillae* in the plastid DNA phylogeny. Furthermore, section *Apodanthi* is sister to the remaining taxa in subgenus *Narcissus,* rather than a clade including sections *Bulbocodii* and *Jonquillae*, as with the plastid phylogenies.

The combined dataset of 77 CDS and nuclear ribosomal sequence consists of 86,875 bp. In the phylogeny (Figure 4-6), subsection *Pseudonarcissi* is again sister to subsection *Narcissus* as in the phylogeny based upon nuclear ribosomal DNA. Moreover, section *Bulbocodii* is sister to section *Jonquillae* as in the phylogenies based upon plastid DNA. Section *Apodanthi* is placed sister to the remaining taxa in subgenus *Narcissus*. Overall the bootstrap support for the phylogeny based upon the combined dataset has strong support (BS ≤95).

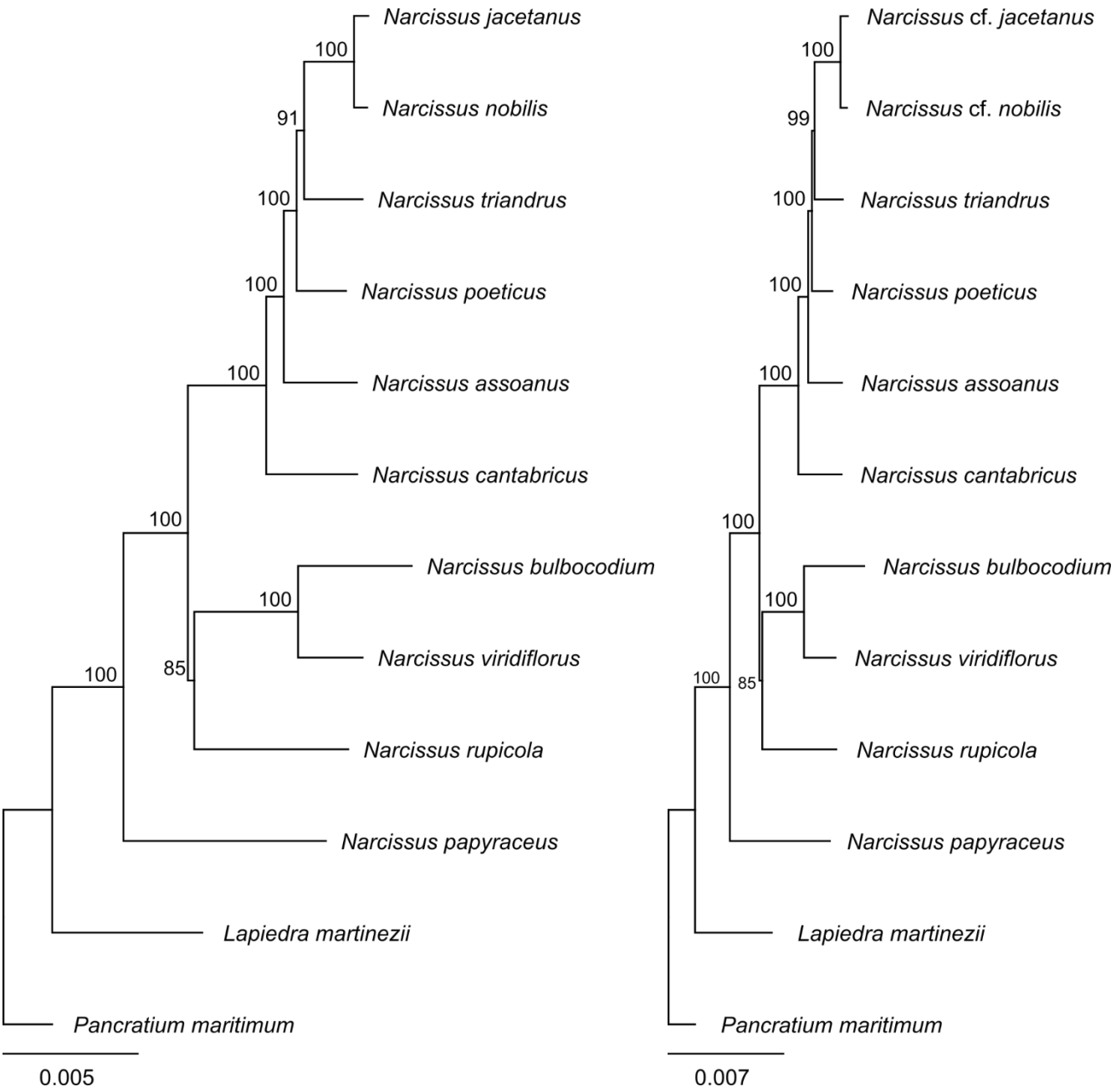


Figure 4-4 RAxML output trees based on the whole plastid genome (left), and 77 cpDNA coding regions (right) for the 12 plastid genomes in this study. Numbers at each node indicate bootstrap support.

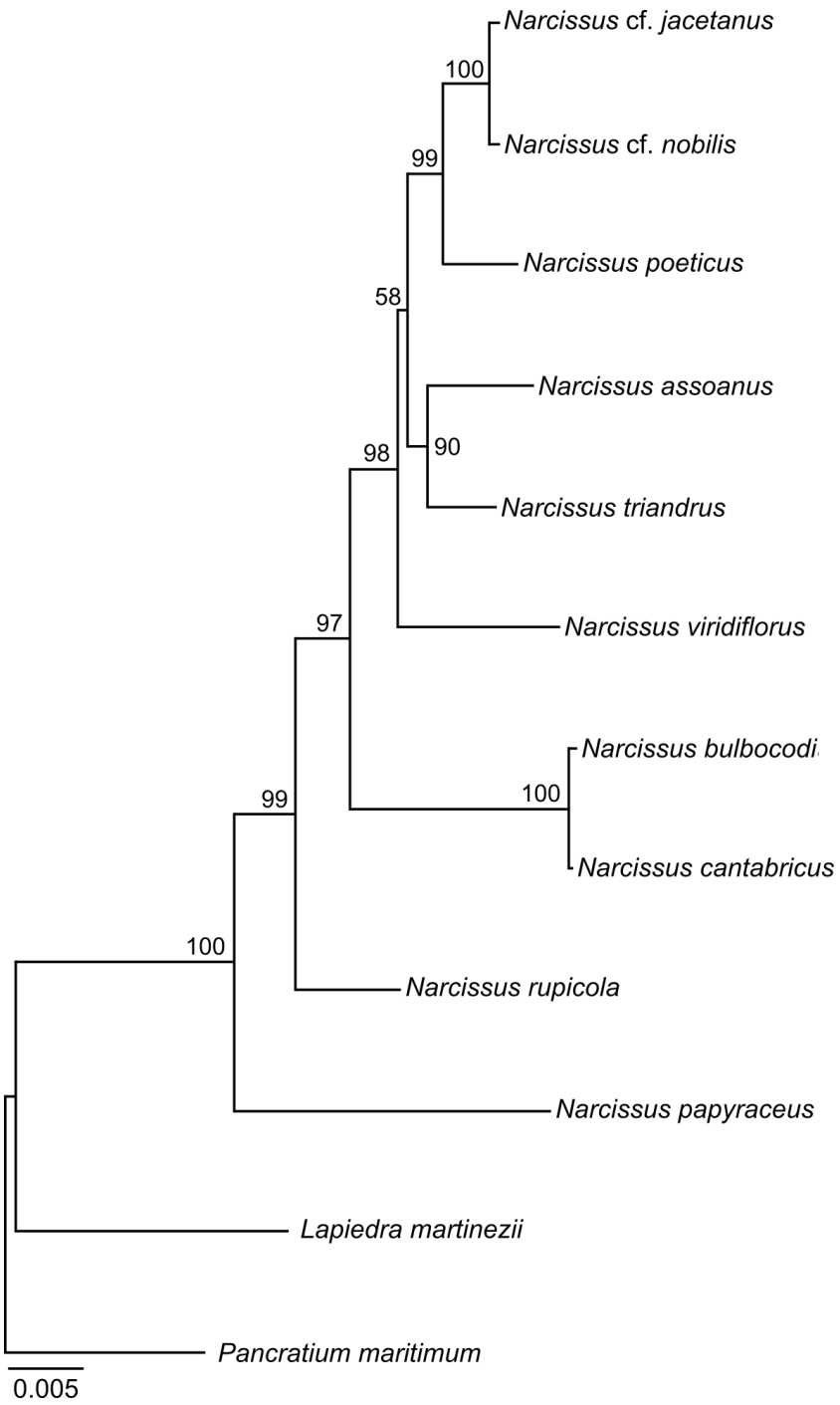


Figure 4-5 RAxML output tree based on the nuclear ribosomal scaffold for the 12 samples within this study. The numbers at each node indicate bootstrap support.

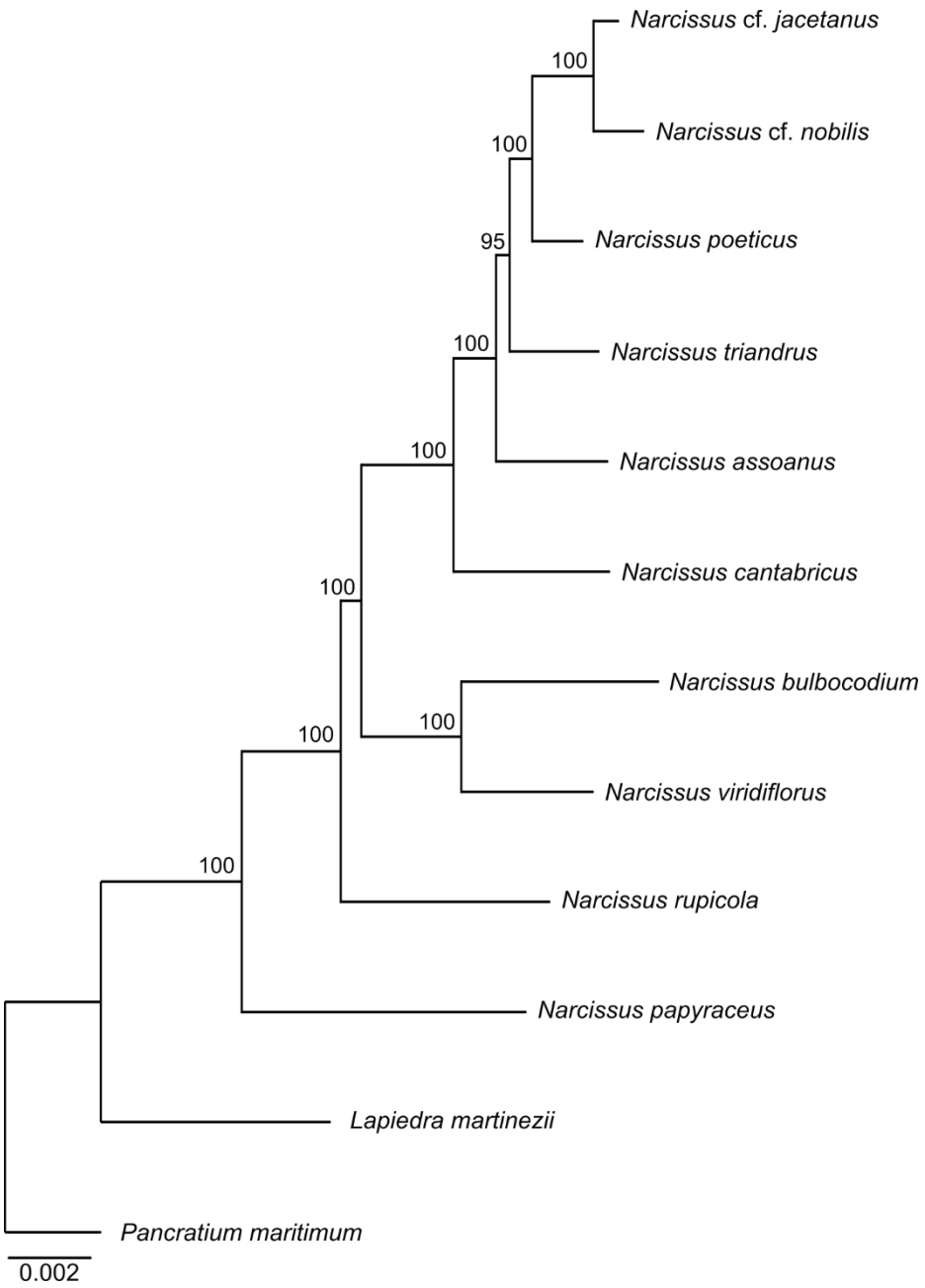


Figure 4-6 RAxML output tree based on the whole plastid genome concatenated with nuclear ribosomal DNA for the 12 samples within this study. The numbers at each node indicate bootstrap support.

4.4.3 Simple sequence repeats

On average of 65 SSRs were found from the 12 plastid genomes within this study (Table 4-4). Six types of SSR were identified across the taxa within this study, for each genome 57-81 loci were identified with a total of 789 SSRs present in the plastid genomes. The most abundant type of SSR was mononucleotide repeats, followed by dinucleotide SSRs, tetranucleotide SSRs before trinucleotide SSRs.

Table 4-4 Simple sequence repeats in the 12 plastid genomes.

Species	**Number of SSR loci**	**Type of nucletotide SSR**						**Repeat location**		
		Mono-	**Di-**	**Tri-**	**Tetra-**	**Penta-**	**Hexa-**	**LSC**	**IR**	**SSC**
N. assoanus	63	42	13	1	6	0	1	52	2	7
N. bulbocodium	63	46	10	1	4	2	0	46	3	11
N. cantabricus	60	37	13	3	7	0	0	43	3	11
N. jacetanus	67	45	12	4	5	1	0	52	2	11
N. nobilis	63	44	11	3	4	0	1	49	2	10
N. papyraceus	58	38	13	1	5	0	1	47	2	7
N. poeticus	68	43	11	7	6	0	1	56	2	8
N. rupicola	74	47	16	4	7	0	0	62	2	8
N. triandrus	65	42	10	3	9	0	1	52	2	9
N. viridiflorus	70	51	14	2	3	0	0	53	2	13
P. maritimum	57	34	12	2	6	3	0	48	3	3
L. martinezii	81	49	15	3	7	4	3	64	3	11

By comparison, in an aligned dataset of 12 plastid genomes 656 SSRs were identified, although only 28 SSRs were found in ≥8 samples, details shown in Table 4-5. Seven of the simple sequence repeats were found in all of the sampled taxa.

Table 4-5 Details of the plastid simple sequence repeats found within ≥8 of the study species. Regions in bold were identified in all of the study taxa.

ID	Motif	Repeat range	Start	End
CP1	(T)	10-12	6756	6767
CP2	(T)	11-15	18047	18061
CP3	**(AT)**	**5**	**21531**	**21540**
CP4	(T)	10-16	28969	28984
CP5	(A)	10-13	30506	30518
CP6	(AT)	5-13	30637	30662
CP7	(A)	10-11	31171	31180
CP8	(TTTA)	3-4	33503	33518
CP9	(A)	10-13	34779	34791
CP10	(T)	10-14	35570	35583
CP11	(AT)	5-8	38640	38655
CP12	(TTTC)	3	44936	44947
CP13	(TTCT)	3	46426	46437
CP14	(T)	10-12	50076	50087
CP15	**(AT)**	**5-6**	**51106**	**51117**
CP16	(AT)	5-6	61920	61931
CP17	**(AATG)**	**3**	**66633**	**66644**
CP18	(T)	10-12	73505	73516
CP19	(A)	10-11	75343	75353
CP20	**(T)**	**10-12**	**75645**	**75655**
CP21	(T)	10	85830	85839
CP22	**(GA)**	**5**	**95325**	**95334**
CP23	(T)	10-11	117500	117510
CP24	(AATA)	3	123979	123990
CP25	(T)	10-11	133420	133499
CP26	(T)	11	134890	134900
CP27	**(A)**	**10**	**135977**	**135986**
CP28	**(TC)**	**5**	**158159**	**158168**

4.5 Discussion

4.5.1 Plastome features

In this chapter nine complete plastomes have been assembled for *Narcissus* species, as well one plastid genome for each *Lapiedra* and *Pancratium,* this is the first comparative study within subfamily Amaryllidoideae. The organisation of the plastomes presented in this study are similar to those available in Asparagales. Inverted repeat regions are typically highly conserved;

however the expansion or contraction of the IR junction boundaries considered the main mechanism for length variation in angiosperms (Raman and Park, 2016). The length variation of the IR found in *N. viridiflorus* was not found in any other daffodil species sampled here, although it was similar to the length variation found in *Lapiedra*. *Narcissus viridiflorus* is a tetraploid, an autopolyploid of *N. jonquilla* (Fernandes, 1975) and was the only known polyploid sampled in this study. However, the addition of further *Narcissus* plastomes is required to reveal the extent of this length variation across the genus.

The *ycf1* pseudogene is common across species within Asparagales, however it is 1,277 bp larger than other species available (Könyves, Bilsborrow, et al., 2018). Within Amaryllidaceae, there are now plastid genomes available for the three subfamilies: Amaryllidoideae (*Narcissus*, *Pancratium* and *Lapiedra*), Alliodeae (*Allium cepa*; KM088013), and Agapanthoideae (*Agapanthus coddii*; KX790363). The pseudo-*ycf1* gene is 1,038 bp and 1,027 bp in available representatives of Allioideae and Agapanthoideae respectively; similar lengths were observed in *N. viridiflorus* and *L. martinezii*, whereas a pseudo-*ycf1* gene of at least 2,641 bp was observed in the remaining samples of Amarylldoideae.

4.5.2 Variable regions

Two of the six most variable regions identified within the plastid genomes were, *matK* and *ndhF.* Both of these regions have been previously used in phylogenetic study of *Narcissus* (Marques, Aguilar, et al., 2017, Könyves, David, et al., 2019). However, the most variable region found was *ycf1*, high variation in *ycf1* has also been found in many other plant groups (Lu, Li, et al., 2017, Bi, Zhang, et al., 2018, Zhao, Song, et al., 2018), meaning it could be an ideal region to use for phylogenetic study (Dong, Xu, et al., 2015). The regions identified could potentially provide the highly variable sites required for further phylogenetic studies; as *rps18*, *ndhD*, and *accD* were all found to be more variable than *matK*, yet have not been applied to study species boundaries within *Narcissus*. Chapter 2 includes sequence data for a partial region of *ycf1* due to the variation found within this study, however, it was still not able to ascertain the species boundaries between *N. asturiensis* and *N. cyclamineus*. This may be due to sequences of *ycf1* used in this thesis were gathered from already published barcode regions (Dong, Xu, et al., 2015) however, using the plastomes presented in this study means further primers could be designed to explore the species level variation within this gene.

Given the prevalent hybridisation within *Narcissus* (Fernandes, 1975, Marques, Feliner, et al., 2010), plastid-specific SSRs would be useful alongside already developed nuclear genome microsatellites (Hodgins, Stehlik, et al., 2007, Simón, Xavier Picó, et al., 2010, Könyves, David, et al., 2016, Barranco, Simón-Porcar, et al., 2019). The number of SSRs detected ranged from 58 (*N. papyraceus*) to 81 (*L. martinezii*), lower than the 179-193 SSRs detected in species of *Allium* (Xie, Yu, et al., 2019), although similar to the number of SSRs found in studies of *Fritillaria* (Bi, Zhang, et al., 2018, Li, Zhang, et al., 2018). Using a combination of plastid and nuclear SSRs would help to elucidate patterns of seed and pollen flow within the genus. Both plastid and nuclear microsatellites have been used to identify gene flow between two species of *Rosa* in Japan (Nagamitsu, 2017). The extent of hybridisation was revealed in a phylogeographic study of *Fraxinus*, using plastid genome microsatellites, indicating their utility for tracing maternal inheritance (Heuertz, Carnevale, et al., 2006). Seven microsatellite loci were discovered in all of the taxa sampled in this study, due to the coverage of the genus the sampling represents, levels of cross-amplification should therefore be high.

4.5.3 Phylogenies

With Next-Generation sequencing technology becoming increasingly more widespread, it is possible to construct phylogenies with large datasets. The most comprehensive plastid sequence *Narcissus* phylogeny to date is based upon 1285 bp from two gene regions, *matK* and *ndhF* (Marques, Aguilar, et al., 2017). In this study, whole plastomes were used to construct initial phylogenies to observe sectional placement within the genus, as well as nuclear ribosomal data to review conflict.

Section *Apodanthi*, represented in this phylogeny by *N. rupicola,* changes position depending on the data source used (Figure 4-4 to 4-6). The species is either found sister to section *Bulbocodii* and *Jonquillae,* (Whole cpDNA, CDS), or sister to the remaining taxa within subgenus *Narcissus* (nuclear ribosomal DNA, and combined). The placement of section *Apodanthi* is therefore unresolved here, although the sampling is not complete as several sections are unsampled and aside from subsection *Pseudonarcissi* only one taxa was sampled per section. Similar incongruence has been found in previous plastid DNA phylogenies, that show section *Apodanthi* as both sister to section *Bulbocodii* and *Jonquillae* (Santos-Gally, Vargas, et al., 2012, Marques, Aguilar, et al., 2017), and as sister to all the remaining taxa in

subgenus *Narcissus* (Graham and Barrett, 2004). Section *Apodanthi* has also been found as sister to sections *Bulbocodii*, *Braxireon*, *Juncifolii* and subsection *Pseudonarcissi* in a cpDNA tree based upon *matK* and *ndhF* (Könyves, David, et al., 2019). With the addition of mitochondrion and further nuclear DNA from the raw sequencing reads the placement of section *Apodanthi* could stabilize. Genome skimming was used to assemble mitochondrial genomes to resolve phylogenetic relationships in Oleaceae, in contrast with the plastome phylogeny (Van de Paer, Bouchez, et al., 2017).

There is incongruence of the placement of sections *Bulbocodii* and *Meridionalis* between the plastid phylogenies and nuclear ribosomal phylogeny. In the plastid phylogenies, section *Bulbocodii* is sister to section *Jonquillae* however based upon nuclear ribosomal date *Bulbocodii* is sister to *Meridionalis*. Sections *Bulbocodii* and *Meridionalis* encompass the hoop-petticoat daffodils, formerly recognised as section *Bulbocodii* based upon morphological evidence. Molecular studies of hoop-petticoat daffodils have revealed separation of section *Bulbocodii* based upon geographic patterns of variation, without any morphological characters to separate the two sections. The phylogenies presented in this study therefore corroborate with previous studies in revealing molecular patterns that suggest the input hybridisation has had upon the evolution of *Narcissus*. Further sampling of subsection *Pseudonarcissi* taxa will need to be added before drawing conclusions regarding the monophyly shown here, as subsection *Pseudonarcissi* has been shown to be polyphyletic in previous phylogenetic studies (Marques, Aguilar, et al., 2017, Könyves, David, et al., 2019).

In this study, the plastid genomes of nine *Narcissus* species, *Pancratium maritimum*, and *Lapiedra martinezii* were reported from *de novo* sequencing, providing the first comparative analysis of plastid genomes for the subfamily Amaryllidoideae. All of the genomes assembled for *Narcissus* displayed similar length variations, besides the contraction of the IR in *N. viridiflorus*, a similar sized length contraction was also found in one outgroup taxon, *L. martinezii*. The plastid genomes assembled in this study provide rich genetic information for phylogenetic study in the future. Further sampling of *Narcissus* taxa is required before drawing assumptions on the sectional placement however, the plastomes in this study provide the first step in using genomics data for *Narcissus*. The assembly of mitochondrial reads from the genome skim data in this study could also help resolve the placement of section *Apodanthi*.

Furthermore, SSRs identified within the plastid genome could be used alongside currently available SSRs for population level studies into the genus. The sampling chosen in this study encompasses the variation of the whole genus to reflect the development of plastid SSRs, therefore providing a high chance of successful cross amplification.

Given the horticultural importance of *Narcissus*, the data from this study will provide rich genetic information for the development of a molecular means for identifying daffodils in horticulture. This study also provides a basis for a future taxonomic revision of the wild species of the genus; identifying variable regions of the plastid genome and providing the data for the development of a suite of SSRs or the identification of single nucleotide polymorphisms. Underpinning the taxonomy of *Narcissus* is an important step in developing a molecular means for daffodil cultivar identification.

Chapter 5 Daffodil Cultivar Identification

5.1 Introduction

The ornamental horticulture industry was worth over £24 billion to the UK economy in 2017 (Oxford Economics, 2018). Most of this trade is based around named cultivars, with 17,000 species in the RHS Plant Finder encompassing 72,000 named cultivars that are available in the UK (Cubey, 2017). The accurate naming of cultivars and a means for identification is therefore essential for the horticultural trade. The plant names provide the link of the horticultural characteristics and appearance that gardeners and plant breeders would expect from their plants. Furthermore, certain cultivars can demand premium prices, for example due to rarity or challenge in growing, it is therefore essential that breeders can accurately name their cultivars.

Cultivars are typically identified using morphological data, this requires meticulous study and comparisons of plant specimens usually by expert groups and associations with many years of experience. Current identification methods can be impractical in groups of plants were there are large numbers of cultivars such as dahlias, with over 48,000 cultivars (McClaren, 2009), or daffodils with 30,000 (Kington, 2008, Daffseek, 2018), as there are often not enough characters to uniquely identify each cultivar. This problem is compounded by seasonality, as diagnostic characters are observed during the growing season and therefore identification is unfeasible outside of this period. For example, daffodils are identified using floral characters however, they are largely traded as bulbs therefore identification cannot be confirmed until later in the growing season.

Daffodils are widely planted as spring ornamentals and have been in cultivation for centuries with one of the earliest illustrations of horticultural daffodils from the 17th Century (van de Passe, 1614), although daffodils are known to have been in cultivation before this time (Coates, 1956). Many cultivars have been in cultivation so long that the origins, pedigree and original hybridiser are not known. Despite the historic importance, intensive breeding programmes over the last century in particular have given rise to 30,000 registered daffodil cultivar names (Kington, 2008, Daffseek, 2018), of which approximately 2,000 are available to purchase in the UK (Cubey, 2017). The sale of daffodil cultivars is worth approximately £45 million to the UK economy (MarketInsider, 2015), with the UK being the second largest producer of cultivated daffodils after the Netherlands (Hanks, 2002a).

New daffodil cultivars are formerly registered with the Royal Horticultural Society which requires the plant breeder to provide diagnostic characters. Registration is becoming an increasingly difficult challenge given the number of registered cultivars and with new cultivars being introduced every year. New cultivars are introduced each year through the continuation of breeding programmes that introduce wild sourced material to horticulture. Due to the continual introduction of wild species, daffodil cultivars must not be studied in isolation of their wild progenitor species. Representatives from the entire genus have been used in breeding programmes to produce variation we see in daffodil cultivars today however, trumpet daffodils have contributed more than any other to daffodil cultivar breeding (Rivera Nuñez, Obón De Castro, et al., 2003).

The 30,000 cultivars are divided into a horticultural classification including twelve Divisions, based on either floral characters, measurements, or progenitor species (Table 1-5). This horticultural classification is artificial and entirely separate from that of wild species and is primarily required for convenience when registering, trading, and displaying daffodil cultivars at horticultural shows (Kington, 2008). It is therefore not expected that these Divisions will be recoverable using molecular data except where a Division is defined by a common parent species (e.g. Division 8, Tazetta daffodil cultivars).

DNA fingerprinting methods have been widely applied due to their suitability to research questions identifying variation within species or closely related species. These techniques are also particularly valuable in the identification of clonally propagated cultivars such as potatoes and daffodils as resolution can extend to distinguishing individual genotypes. This makes DNA fingerprinting methods useful for identifying cultivated plants, as reviewed by Culham & Grant (1999). Microsatellites have been the most commonly used DNA fingerprinting method over the past two decades, and their utility to cultivar identification has been demonstrated in ornamental crops such as Dianthus (Arens, Esselink, et al., 2009), *Paeonia* (Gilmore, Bassil, et al., 2013), and *Pelargonium* (Peltier, Becher, et al., 2000). A set of four microsatellite loci were able to ascertain the identification of 24 *Anthurium* cultivars (Wang and Chuang, 2013). Microsatellites have also proven to be a useful tool in economically important crops for cultivar and pedigree identification (Galli, Halasz, et al., 2005, Miah, Rafii, et al., 2013, Ben Ayed, Ben Hassen, et al., 2016). This is despite the drawbacks of using microsatellites due to; difficulties

amplifying regions of the genome with high mutation rates, and species specificity of priming sequences (Sinama, Dubut, et al., 2011). These issues can be overcome by creating larger libraries of data that are becoming increasingly common with Next-Generation Sequencing techniques, the rapid uptake of NGS technology has provided whole genome databases in ornamental plants (Yagi, 2018). However, these genomic resources are currently not widely available in non-model organisms.

To date molecular tools have not been used for daffodil cultivar identification, although the cross amplification of microsatellites have been tested to daffodil cultivars (Hodgins, Stehlik, et al., 2007). A molecular methodology for cultivar identification would allow plants to be quickly characterised by breeders at any stage of the growing season, and would help to identify diagnostic characters for registering cultivars. A molecular tool for identification could also aid in the identification of synonyms, mislabelled, and mis-sold stocks.

In this study we apply microsatellite markers developed for hoop-petticoat daffodils (Könyves, David, & Culham, 2016) to test their utility for identifying daffodil cultivars across all horticultural Divisions.

5.2 Objectives

The aim of this chapter is to demonstrate whether microsatellite markers can provide a molecular tool for the identification of daffodil cultivars. Thus, the objectives for this chapter are:

1. To identify the provenance and pedigree of daffodil cultivars from all horticultural Divisions, using plastid DNA haplotypes and SSR data.
2. To identify if SSR data can be used to successfully discriminate amongst a representative sample of commercially sourced daffodil cultivars.

5.3 Material and Methods

5.3.1 Plant material

Leaf material of 114 cultivars was sourced through commercial suppliers in the UK Leaf samples of an additional 22 leaf samples were sourced through collaborators at RHS Wisley, and a National Collection® of daffodils (Croft16). Details of the sampled cultivars, including their

parentage is shown in Table 5-1. For each horticultural Division up to 11 cultivars were collected, aside from Division 10 were 24 cultivars was sourced. For every other Division up to 11 cultivars were obtained. For four cultivars: *N.* 'Julia Jane', *N.* 'Spoirot', *N.* 'Golden Bells', and *N.* 'Tarlatan' accessions were sourced from different suppliers. Cultivar leaves were collected and stored in silica gel.

Table 5-1 Details of sampled daffodil cultivars (Kington, 2008). D = Division; A = Cultivar code corresponding to analysis; M = *matK* haplotype; K = DAPC cluster. For cultivars with more than one sample, the supplier code is hyphenated. Divisions of parentage shown in brackets.

D	Cultivar	A	Parentage	M	K
1	*N.* 'Cedric Morris'	1	Selection of *N. minor* (13)	H9	1
1	*N.* 'Cristobal'	2	*N.* 'My Love' (2) × open pollinated	H1	1
1	*N.* 'Dutch Master'	3	Not known	H1	1
1	*N.* 'Golden Spur'	4	Not known	H1	1
1	*N.* 'Immaculate'	5	[(*N.* 'Green Valley' (2) × *N.* 'Tetecta' (2))] × [(*N.* 'Tudor Minstrel' (2) × *N.* 'Amorette' (2))]	H1	1
1	*N.* 'Jamboree'	6	*N.* 'Kingscourt' (1) × *N.* 'Chemawa' (2)	H1	1
1	*N.* 'Madame Plemp'	7	*N.* 'Empress' (1) × *N.* 'Madame de Graaff' (1)	H1	2
1	*N.* 'Magnificence'	8	*N.* 'Maximus' (1) × *N.* 'Leedsii' (*N. hispanicus* var. *spurius* × unknown)	H1	1
1	*N.* 'Memento'	9	*N.* 'Gloriola' (2) × *N.* 'Rima' (1)	H1	1
1	*N.* 'Midas Touch'	10	*N.* 'Camelot' (2) × *N.* 'Arctic Gold' (1)	H1	1
1	*N.* 'Rame Head'	11	*N.* 'Kingscourt' (1) × open pollinated	H1	1
1	*N.* 'Viking'	12	*N.* 'Goldcourt' (1) × *N.* 'Kingscourt' (1)	H8	1
1	*N.* 'Welsh Rugby Union'	13	*N.* 'Argosy' (1) × [*N.* 'Saint Keverne' (2) × *N.* 'Malvern City' (1)]	H1	1
2	*N.* 'Amber Castle'	14	*N.* 'Camelot' (2) × *N.* 'Daydream' (2)	H1	1
2	*N.* 'Balvenie'	15	*N.* 'Merry Widow' (2) × [*N.* 'Salmon Trout' (2) × (*N.* 'Rosedew' (2) × *N.* 'Rose Caprice' (2))]	H1	1
2	*N.* 'Binkie	16	Not available	H1	1
2	*N.* 'Cameo King'	17	*N.* 'Cool Crystal' (3) × *N.* 'Immaculate' (2)	H8	1
2	*N.* 'Carlton	18	Not known	H1	1
2	*N.* 'Culmination'	19	*N.* 'Romance' (2) × *N.* 'Cool Flame' (2)	H1	1
2	*N.* 'Fair Prospect'	20	*N.* 'Infatuation' (2) × *N.* 'Debutante' (2)	H1	1
2	*N.* 'Orange Tint'	21	*N.* 'Prologue' (1) × *N.* 'Johann Strauss' (2)	H1	1
2	*N.* 'Saint Keverne'	22	(*N.* 'Royalist' (1) × unknown) × unknown	H1	1
2	*N.* 'Strines'	23	*N.* 'Golden Torch' (2) × *N.* 'Kingscourt' (1)	H1	1
3	*N.* 'Albatross'	24	*N.* 'Ornatus' (9) × *N.* 'Empress' (1)	H1	1
3	*N.* 'Cool Crystal'	25	*N.* 'Chinese White' (3) hybrid	H1	1
3	*N.* 'Dallas'	26	*N.* 'Cushendall' (3) × *N.* 'Silver Salver' (3)	H1	1
3	*N.* 'Edward Buxton'	27	Not known	H1	1
3	*N.* 'Fairlawns'	28	[*N.* 'Corofin' (3) × N. 'Kimmeridge' (3)] × *N.* 'Merlin' (3)	H1	1
3	*N.* 'Halleys Comet'	29	*N.* 'Verona' (3) × *N.* 'Thoresby' (3)	H1	1
3	*N.* 'La Riante'	30	Not known	H1	2
3	*N.* 'Painted Desert'	31	*N.* 'Old Satin' (2) × *N.* 'Altruist' (3)	H1	1
3	*N.* 'Portrush'	32	*N.* 'Silver Coin' (3) × *N.* 'Crimson Braid' (3)	H1	1
3	*N.* 'Queen of the North'	33	Not available	H1	1
3	*N.* 'Seraglio'	34	*N.* 'Mozart' (2) × *N.* 'Gallipoli' (2)	H1	1

Table 5-1 (cont.) Details of sampled daffodil cultivars (Kington, 2008). D = Division; A = Cultivar code corresponding to analysis; M = *matK* haplotype; K = DAPC cluster. For cultivars with more than one sample, the supplier code is hyphenated. Divisions of parentage shown in brackets.

D	Cultivar	A	Parentage	M	K
3	*N.* 'White Lady'	35	*N.* 'Ornatus' (9) × *N. triandrus* subsp. *pallidulus* (13)/*N. moschatus* (13)	H1	1
4	*N.* 'Acropolis'	36	*N.* 'Falaise' (4) × *N.* 'Limerick' (3)	H1	1
4	*N.* 'Angkor'	37	*N.* 'Gay Time' (4) × *N.* 'Daydream' (2)	H1	1
4	*N.* 'Blushing Maiden'	38	*N.* 'Pink Chiffon' (4) × *N.* 'Accent' (2)	H1	1
4	*N.* 'Golden Bear'	39	*N.* 'Smokey Bear' (4) × *N.* 'Sportsman' (2)	H1	1
4	*N.* 'Merry Meet'	40	*N.* 'Falaise' (4) × *N.* 'King's Ransom' (1)	H1	1
4	*N. poeticus* 'Flore Pleno'	41	Selection of *N. poeticus* subsp. *recurvus* (13)	H1	2
4	*N.* 'Rip Van Winkel'	42	*N. pumilus* (13) variant	H9	1
4	*N.* 'Sir Winston Churchill'	43	*N.* 'Geranium' (8) sport	H1	2
4	*N.* 'Sugar Loaf'	44	*N.* 'Pink Chiffon' (4) × *N.* 'Accent' (2)	H1	1
4	*N.* 'Van Sion'	45	Syn. *N.* 'Telamonius Plenus'	H1	1
4	*N.* 'Wheal Bush'	46	*N.* 'Saint Keverne' (2) × *N.* 'Tamar Fire' (4)	H1	1
4	*N.* 'Yellow Cheerfulness'	47	*N.* 'Cheerfulness' (4) sport	H1	1
5	*N.* 'Angel's Whisper'	48	*N. triandrus* (13) × *N. fernandesii* (13)	H2	1
5	*N.* 'Chipper'	49	[*N.* 'Polindra' (2) × *N.* 'Tunis' (2)] × *N. triandrus* (13)	H1	1
5	*N.* 'Fairy Chimes'	50	*N. jonquilla* (13) × *N. triandrus* (13)	H3	1
5	*N.* 'Harmony Bells'	51	*N.* 'Whiteley Gem' (2) × *N. triandrus* (13)	H1	1
5	*N.* 'Lemon Drops'	52	*N.* 'Fortune' (2) × *N. triandrus* (13)	H1	1
5	*N.* 'Mission Bells'	53	*N.* 'Silver Bells' (5) × open pollinated	H1	1
5	*N.* 'Puppet'	54	*N.* 'Narvik' (2) × *N. triandrus* var. *concolor* (13)	H1	1
5	*N.* 'Stint'	55	[*N.* 'Fortune's Sun' (2) × *N.* 'Cheerio' (2)] × *N. triandrus* (13)	H1	1
5	*N.* 'Thalia'	56	*N. triandrus* (13) × 2W-W seedling	H2	1
5	*N.* 'Tuesday's Child'	57	*N.* 'Interim' (2) × *N. triandrus* var. *loiseleurii* (13)	H1	1
6	*N.* 'Beryl'	58	*N.* 'Chaucer' (9) × *N. cyclamineus* (13)	H1	1
6	*N.* 'Cornet'	59	*N. cyclamineus* (13) seedling	H1	1
6	*N.* 'Crofty'	60	*N.* 'Perimeter' (3) × *N. cyclamineus* (13)	H1	1
6	*N.* 'February Gold'	61	*N.* 'Golden Spur' (1) × *N. cyclamineus* (13)	H1	1
6	*N.* 'Jetfire'	62	[(*N.* 'Market Merry' (3) × *N.* 'Carbineer' (2)) × *N.* 'Armada' (2)] × *N. cyclamineus* (13)	H1	1
6	*N.* 'Lily May'	63	*N.* 'Foundling' (6) × *N.* 'Raspberry Ring' (2)	H9	1
6	*N.* 'Rapture'	64	*N.* 'Nazareth' (2) × *N. cyclamineus* (13)	H1	1
6	*N.* 'Sissy'	65	*N.* 'Perimeter' (3) × *N. cyclamineus* (13)	H1	1
6	*N.* 'Snipe'	66	*N.* 'W.P Milner' (1) × *N. cyclamineus* (13)	H1	1
6	*N.* 'Trewirgie'	67	Not known	H1	2
6	*N.* 'Turncoat'	68	*N.* 'Richhill' (2) × *N.* 'Foundling' (6)	H1	1
7	*N.* 'Boscastle'	69	*N.* 'Aircastle' (3) × *N. jonquilla* (13)	H1	1

Table 5-1 (cont.) Details of sampled daffodil cultivars (Kington, 2008). D = Division; A = Cultivar code corresponding to analysis; M = *matK* haplotype; K = DAPC cluster. For cultivars with more than one sample, the supplier code is hyphenated. Divisions of parentage shown in brackets.

D	Cultivar	A	Parentage	M	K
7	*N.* 'Canary'	70	*N.* 'Daydream' (2)× *N. jonquilla* (13)	H1	1
7	*N.* 'Crill'	71	*N.* 'Dateline' (3) × *N. jonquilla* (13)	H1	1
7	*N.* 'Flycatcher'	72	[*N.* 'Playboy' (2) × *N.* 'Firecracker' (2)] × *N. assoanus* (13)	H14	1
7	*N.* 'Limequilla'	73	*N.* 'Lime Chiffon' (1) × *N. jonquilla* (13)	H1	1
7	*N.* 'Pacific Coast'	74	*N.* 'Minnow' sport	H1	1
7	*N.* 'Pueblo'	75	*N.* 'Binkie' (2) × *N. jonquilla* (13)	H1	1
7	*N.* 'Sabrosa'	76	*N. jonquilla* (13) × *N. rupicola* subsp. *watieri* (13)	H14	1
7	*N.* 'Step Forward'	77	*N.* 'Quick Step' (7) × *N.* 'Daydream' (2)	H1	1
7	*N.* 'Sweet Blanche'	78	*N.* 'Grand Prospect' (2) × *N. jonquilla* (13)	H1	1
8	*N.* 'Brentswood'	79	Not known	H13	1
8	*N.* 'Canaliculatus'	80	Not known	H12	1
8	*N.* 'Chinita'	81	*N.* 'Chaucer' (9) × *N.* 'Jaune a Merveille' (8)	H1	1
8	*N.* 'Grand Primo Citroniere'	82	Not known	H13	1
8	*N.* 'Jamage'	83	*N.* 'Grand Primo' (8) × *N.* 'Green Island' (2)	H13	1
8	*N.* 'Killara'	84	*N.* 'Grand Monarque' (8) × *N.* 'Empress' (1)	H13	1
8	*N.* 'Martinette'	85	*N.* 'Matador' (8) × *N. jonquilla* (13)	H12	1
8	*N.* 'Mike Pollock'	86	*N.* 'Matador' (8) × *N. jonquilla* (13)	H12	1
8	*N.* 'Pollys Pearl'	87	Selection from *N.* 'White Pearl' (8)	H13	1
8	*N.* 'St. Agnes'	88	*N.* 'Chaucer' (9) × *N. tazetta* (13)	H1	1
9	*N.* 'Cassandra'	89	*N.* 'Ornatus' (9) × *N. radiiflorus* var. *poetarum* (13)	H1	2
9	*N.* 'Hexameter'	90	*N.* 'Raeburn' (9) × *N.* 'Dactyl' (9)	H1	1
9	*N.* 'Horace'	91	*N.* 'Ornatus' (9) × *N. radiiflorus* var. *poetarum* (13)	H1	2
9	*N.* 'Kabani'	92	*N.* 'Woodland Star' (3) × *N.* 'Lady Serena' (9)	H1	1
9	*N.* 'Merthan'	93	*N.* 'Corofin' (3) × *N.* 'Lady Serena' (9)	H1	1
9	*N.* 'Milan'	94	Not known	H1	1
9	*N.* 'Penjerrick'	95	*N.* 'Woodland Star' (3) × *N.* 'Lady Serena' (9)	H1	1
9	*N.* 'Praecox'	96	*N. poeticus* (13) seedling	H1	1
9	*N.* 'Sonata'	97	Not known	H1	1
10	*N.* 'Apollo Gold'	98	*N. bulbocodium* var. *conspicuus* (13) × *N. jonquilla*(13)	H5	2
10	*N.* 'Atlas Gold'	99	Selection of *N. romieuxii* (13)	H5	3
10	*N.* 'Buttercream'	100	Seedling 8-17 × *N.* 'Peppermint' (10)	H10	3
10	*N.* 'Casual Elegance'	101	Seedling 8-17 × *N.* 'Peppermint' (10)	H10	3
10	*N.* 'China Gold'	102	(*N. bulbocodium* (13) × unknown) × *N. rupicola* (13)	H10	2
10	*N.* 'Classic Gold'	103	*N. bulbocodium* var. *conspicuus* (13) × *N. rupicola* (13)	H10	2

Table 5-1 (cont.) Details of sampled daffodil cultivars (Kington, 2008). D = Division; A = Cultivar code corresponding to analysis; M = *matK* haplotype; K = DAPC cluster. For cultivars with more than one sample, the supplier code is hyphenated. Divisions of parentage shown in brackets.

D	Cultivar	A	Parentage	M	K
10	*N.* 'Diamond Ring'	104	Selection of *N. bulbocodium* subsp. *obesus* (13)	H7	2
10	*N.* 'Garden Romance'	105	Seedling 8-7 × *N. cantabricus* subsp. *monophyllus*(13)	H4	3
10	*N.* 'Golden Bells'-MB	106	Selection of *N. bulbocodium* (13)	H7	2
10	*N.* 'Golden Bells'-PM	107		H7	2
10	*N.* 'Golden Bells'-W	108		H7	2
10	*N.* 'Julia Jane' -BB	109	Selection of *N. romieuxii* (13)	H5	3
10	*N.* 'Julia Jane'-MB	110		H5	3
10	*N.* 'Julia Jane'-PM	111		H5	3
10	*N.* 'Kholmes'	112	*N. bulbocodium* var. *conspicuus* (13) × *N. cantabricus* subsp. *foliosus* (13)	H10	3
10	*N.* 'Little Soldier'	113	*N. bulbocodium* subsp. *obesus* (13) × *N.* 'Chemawa' (2)	H7	2
10	*N.* 'Matchmaker'	114	Seedling 10-10 × *N.* 'Peppermint' (10)	H10	3
10	*N.* 'Oxford Gold'	115	(*N. bulbocodium* (13) × unknown) × *N. jonquilla* (13)	H10	2
10	*N.* 'Solveigs Song'	116	*N. bulbocodium.* var. *tenuifolius* (13) × *N. triandrus* subsp. *pallidulus* (13)	H7	2
10	*N.* 'Spoirot' (JB)	117	*N. bulbocodium* var. *conspicuus* (13) × *N. cantabricus* subsp. *foliosus*(13)	H10	3
10	*N.* 'Spoirot' (JP)	118		H10	3
10	*N.* 'Spoirot' (MB)	119		H10	3
10	*N.* 'Spring Medley'	120	Seedling 10-10 × *N.* 'Peppermint' (10)	H10	3
10	*N.* 'Sugar and Cream'	121	Seedling 10-10 × *N.* 'Peppermint' (10)	H10	3
10	*N.* 'Sweet Angel'	122	Seedling 8-17 × *N.* 'Peppermint' (10)	H10	3
10	*N.* 'Symphony Gold'	123	Seedling 8-17 × *N.* 'Peppermint' (10)	H10	3
10	*N.* 'Taffeta'	124	*N. cantabricus* subsp. *foliosus* (13) × *N. romieuxii* (13)	H6	3
10	*N.* 'Tarlatan' (JB)	125	*N. cantabricus* subsp. *foliosus* (13) × *N. romieuxii* (13)	H6	3
10	*N.* 'Tarlatan' (LS)	126		H6	3
10	*N.* 'Treble Chance'	127	Selection of *N. romieuxii* (13)	H5	3
10	*N.* 'Winter Beauty'	128	Seedling 8-18 × *N. cantabricus* subsp. *monophyllus* (13)	H10	3
11	*N.* 'Arwenack'	129	*N.* 'Brandaris' (11) × *N.* 'Paricutin' (2)	H1	1
11	*N.* 'Bosinney'	130	*N.* 'Perlax' (11) × *N.* 'Audubon' (2)	H1	1
11	*N.* 'Fashion'	131	*N.* 'Nippon' (11) hybrid	H1	1
11	*N.* 'Gold Cache'	132	*N.* 'Honeybird' (1) × open pollinated	H1	1
11	*N.* 'Mondragon'	133	Seedling?	H1	1
11	*N.* 'Papillon Blanc'	134	Not known	H1	1
11	*N.* 'Phantom'	135	*N.* 'Accent' (2) × [(*N.* 'Wild Rose' (2) × *N.* 'Hillbilly' (11))]	H2	1
11	*N.* 'Reverse Image'	136	*N.* 'Split' (11) × *N.* 'Salem' (2)	H1	1

5.3.2 DNA extraction

Total genomic DNA was extracted from 10mg of dried leaf material using a modified cetyltrimethylammonium bromide (CTAB) protocol (Doyle and Doyle, 1987). The quality of DNA extractions was visualised using 0.8% agarose gel electrophoresis in 1× TAE buffer stained with GelRed® (Biotium, Inc. Fremont, CA, USA). Quantity and estimated fragment size of the DNA extractions were measured by comparison with Hyperladder 1kb™ (Bioline Reagents Limited, London, UK). Furthermore, a NanoDrop Lite (Thermo Fisher Scientific Inc., Waltham, MA, USA) was used to check concentration and purity. Aliquots of each sample were diluted to 5ng/µl for subsequent PCR amplification.

5.3.3 DNA Sequencing & Fingerprinting

A part of the *matK* gene was amplified for all 136 cultivars using primers *matK*2.1 and *matK*5 (Ford, Ayres, et al., 2009). PCR conditions for the amplification of the *matK* region are provided in Chapter 2. The concentration and size of the PCR products was estimated using 1% agarose gel electrophoresis in 1×TAE with Hyperladder 1kb™ (Bioline Reagents Limited, London, UK) used as a reference. Sequencing of amplicons was carried out by Eurofin Genomics (Cologne, DE) in forward and reverse directions with the sequencing primers. Consensus sequences were assembled in Seqman Pro version 13.0.422 (DNAStar, Madison US). The *matK* sequences generated for daffodil cultivars in this chapter, were compiled with *matK* data from wild species generated in Chapter 2 and from Könyves, David, and Culham (2019).

PCR amplification of microsatellites was performed following the protocol in Könyves, David & Culham (2016). All 11 variable markers were used to genotype the 136 cultivar accessions, see Table 5-2. PCR conditions for amplification of the microsatellites are provided in Chapter 2. Amplicons were visualised using 2% agarose gel electrophoresis in 1×TAE buffer. Approximate size was measured using Hyperladder 100bp™ (Bioline Reagents Limited, London, UK). The amplicons were multiplexed according to Könyves, David & Culham (2016), shown in Table 5-2 and capillary electrophoresis undertaken by Source Bioscience (Nottingham, UK).

Table 5-2 Details of the microsatellites used in this study, information gathered from Könyves, David, et al. (2016).

Locus	Primer sequence	Repeat motif	Allele size range	T_a(°C)	Dye	Multiplex Group
NSB14	F: TGTGTAAGCATACTAACGTTTCG R: AAAAGAGCACCAAGGATGAA	$(ATGT)_{14}$	147-221	48	6-FAM	1
NSB52	F: CAATGGTGGAGCCTCTAATAGC R: TGTCATTCTTTACTTTGTTCTCATTCA	$(GT)_{10}$	117-138	59	PET	3
NSB73	F: GGAGAGGAGTGAGTGGAGTGA R: CAGGCTGTTCAACTATCTTGC	$(CTT)_8$	142-221	60	6-FAM	3
NSB113	F: TTGTGATAAATAAAGGTGCAACTCA R: CATTGCCCGTGATAAGCTCT	(AGT)6	78-123	54	6-FAM	2
NSB122	F: CAAAGTGTTTGTGAATTGCTTC R: GCAATGAGGAGCTTATGAATTAAC	$(AC)_9$	169-198	59	PET	2
NSB143	F: CTGTTTCTTTGTTCTGCACATT R: TCCCAAAATTGCTTCTGAGC	$(GAA)_7$	244-269	59	NED	3
NSB182	F: TTGTATTATACGTTGTTCTGGGGA R: GAGATGCTGACACGCAAACT	$(AC)_{12}$	115-121	58	VIC	1
NSB232	F: CTCCACTTTGGTTGAATCCC R: GACTACCTCCTATTCTAAATGCCA	$(CT)_9$	110-118	63	VIC	2
NSB253	F: GAGGATTACTGTAGCCAATTCCA R: GGACTACAAGATGGCTTCCA	$(GTT)_{11}$	100-140	56	NED	2
NSB263	F: CGAAGGAGGAGTCTTGGAAA R: GAGCAAACTCCTGGCTGAAG	$(GAA)_{11}$	94-131	60	VIC	3
NSB272	F: GGTTCTGCCGATGGACTAAT R: TTATCACATCCAACGGTTTGC	$(CT)_9$	122-146	60	NED	1

The electropherograms were analysed using the microsatellite plug-in of Geneious v.10.2.2 (Kearse, Moir, et al., 2012). Automatic scoring was completed in Geneious and checked manually to avoid overestimation or incorrect peak calling. Alleles were scored according to MANUAL 8 scoring protocol (Pfeiffer, Roschanski, et al., 2011), described in Chapter 2.

5.3.4 Data Analysis

The combined *matK* sequences were aligned in AliView 1.23 (Larsson, 2014) using the MUSCLE 3.8.425 algorithm (Edgar, 2004). A statistical parsimony network was constructed from the *matK* alignment using TCS 1.21 (Clement, Posada, et al., 2000) in order to observe which wild species plastid DNA haplotypes were present in commercially sold daffodils, and if possible to trace the maternal parent of hybrids.

A binary dataset was created from the microsatellite data, to account for polyploids which can have multiple copies of observed alleles, using the presence (1) or absence (0) for each of the

observed alleles in the 136 genotyped samples. Missing data were indicated as "-1". Basic diversity estimates, heterozygosity (H_o) and complement of Simpson's diversity were calculated to estimate the allelic diversity and identification power of the microsatellite loci.

A Provesti's distance matrix was calculated from the binary dataset using the ade4 ver. 1.7-11 (Dray & Dufour, 2007) *R* package, as Provesti's distance is able to handle missing data. Missing data are treated as the average allele frequency. A neighbor-joining tree was constructed from the distance matrix using ape ver. 5.1 (Paradis and Schliep, 2019), in order to establish the relationship amongst cultivars.

Discriminant analysis of principal components (DAPC) was undertaken from the binary dataset using adegenet ver. 2.1.1 (Jombart, 2008), in order to observe groupings of cultivars with horticultural Division used in place of population. All of the principal components were retained and the optimal number of clusters identified by finding the smallest number of clusters, beyond which increasing the number did not lead to a decrease in the Bayesian Information Criterion (BIC) value (Figure 5-1). All k-values from 2 to 4 were selected and the DAPC plotted to identify the optimal number of clusters, as it was not clear from the BIC values. The DAPC clusters were subsequently overlaid onto the NJ tree to observe how they related to genetic distance.

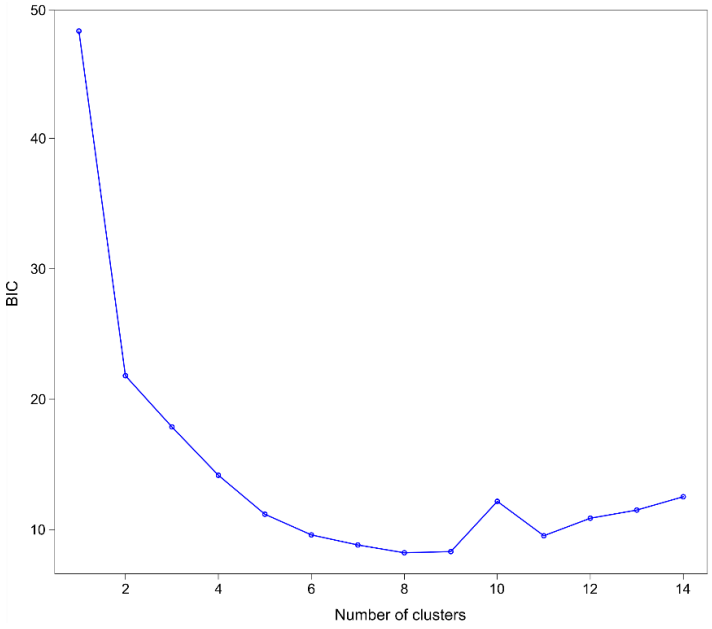


Figure 5-1 Value of Bayesian Information Criterion versus number of clusters. Used to identify the optimal number of clusters.

5.4 Results

5.4.1 Plastid DNA variation

The 136 daffodil cultivars are grouped into 14 haplotypes based on *matK* sequences (Figure 5-2). Exactly which cultivar relates to each of the observed haplotypes are shown in Table 5-1. The most populous haplotype (H1) accounts for 86 of the 136 (63%) cultivars sampled, including cultivars from all but Division 10. The Division 10 cultivars form two separate haplotype groups (H4-7 and H10) separated by a minimum of 16 base pairs. Division 8 daffodil cultivars are distributed amongst two haplotypes (H1, H13), two separated by a minimum of 11 bp whilst two cultivars: *N.* 'Saint Agnes', and *N.* 'Hexameter' cluster in the most populous haplotype (H1). Some cultivars of Divisions 5 and 7 separately form two clusters, however the majority of samples for these two Divisions are found in H1. One Division 11 cultivar, *N.* 'Phantom', clusters with Division 5 cultivars in H2. Haplotype H9 corresponds to Divisions 1, 3, and 5.

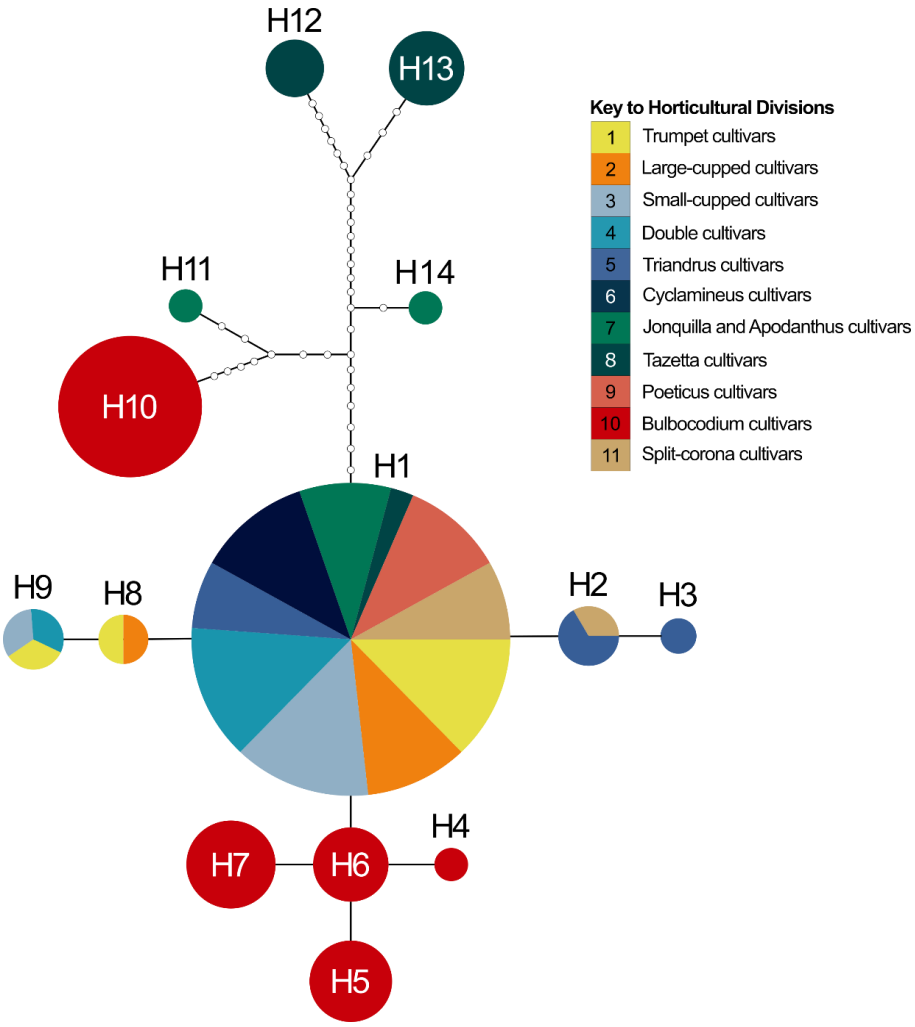


Figure 5-2 Haplotype network of daffodil cultivar *matK* sequences. Haplotypes are coloured according to proportion observed for each horticultural Division. Open circles are inferred haplotypes, length of the line is not indicative of genetic distance. Haplotype numbers are referenced in text.

The *matK* sequence data of all 136 cultivars was compiled with available sequences for wild species, a comparison of the haplotypes found in the cultivar data (Figure 5-2) with the haplotypes found in Chapter 2 are presented in Table 5-3. Only two haplotypes were found in the cultivar data that do not correspond to available data for wild species, H3 and H4. The most populous haplotype found in cultivated daffodils (H1) corresponds to *Narcissus* section *Narcissus*. Haplotypes H3 and H4 comprises only one cultivar each. Two other haplotypes H2 and H9 correspond to subsection *Pseudonarcissi*. Haplotypes H12 and H13 correspond to sections *Angustifolii* and *Tazettae* respectively.

Table 5-3 Haplotype assignment of *matK* sequences for the daffodil cultivars, with comparison to wild species haplotypes presented in Chapter 2.

Cultivar Haplotype	**Species Haplotype**	**Section/Subsection**	**Species**
H1	H1	*Narcissus*	*N.* cf. *pseudonarcissus*, *N. poeticus*
H2	H34	*Pseudonarcissi*	*N. alpestris*
H3	H4	*Ganymedes*	*N. triandrus*
H4	No match	-	-
H5	H40	*Meridionalis*	*N. cantabricus*, *N. romieuxii*
H6	H39	*Meridionalis*	*N. cantabricus*
H7	H38	*Meridionalis*	*N. obesus*
H8	No match	-	-
H9	H5	*Pseudonarcissi*	*N. minor* group (incl. *N. cyclamineus*)
H10	H7	*Bulbocodii*	*N. bulbocodium*
H11	H11	*Apodanthi*	*N. rupicola*
H12	H51	*Angustifolii*	*N. elegans*
H13	H49	*Tazettae*	*N. papyraceus*
H14	H15	*Jonquillae*	*N. jonquilla*

5.4.2 Microsatellite variation

The successful cross amplification of microsatellite markers to daffodil cultivars varied from 39-100%. Markers NSB52 and NSB73 had limited transferability at 58% and 39% successful cross amplification respectively. In contrast, four of the markers (NSB113, NSB232, NSB253, and NSB272) transferred to 100% of the daffodil cultivars. The remaining five markers amplified in over >90% of daffodil cultivars sampled here. All microsatellite loci were polymorphic, with 4-11 alleles identified. Allelic diversity was calculated to be between 0.370 and 0.779, and heterozygosity between 0.044 and 0.647 (Table 5-4). Markers NSB52 and NSB73 had the highest allelic diversity, although they had the lowest cross amplification success.

Table 5-4 Microsatellite marker variation in the genotyped daffodil cultivars. H_0=observed heterozygosity.

Marker name	No. genotyped	No. of alleles	Size range	H_o	Allelic diversity
NSB14	133	6	147-199	0.647	0.625
NSB52	78	8	117-131	0.257	0.785
NSB73	53	9	145-172	0.044	0.779
NSB113	136	4	87-96	0.176	0.39
NSB122	127	9	171-199	0.162	0.624
NSB143	131	8	247-268	0.654	0.734
NSB182	126	4	114-120	0.375	0.578
NSB232	136	5	109-117	0.257	0.37
NSB253	136	9	89-125	0.213	0.548
NSB263	135	11	108-138	0.676	0.705
NSB272	136	7	132-144	0.632	0.729

Six of the microsatellite loci were found to have no variation within particular horticultural Divisions, NSB52, NSB73, NSB113, NSB122, NSB232, and NSB253 (Table 5-5). Loci NSB52 and NSB73 have the largest range of allelic diversity; with maximum diversity in Divisions 1, 6, 7, and 8 however, this is influenced by missing data through a failure to cross-amplify. Within Division 9 the marker NSB73 failed to amplify in any of the accessions. Allele diversity in loci NSB122 and NSB253 was highest in Division 10 cultivars.

Table 5-5 Allelic diversity of microsatellite loci per horticultural Division.

	NSB14	NSB52	NSB73	NSB113	NSB122	NSB143	NSB182	NSB232	NSB253	NSB263	NSB272
Division 1	0.51	0.69	1.00	0.26	0.44	0.59	0.54	0.45	0.41	0.64	0.65
Division 2	0.60	0.60	0.00	0.00	0.17	0.53	0.51	0.42	0.00	0.72	0.68
Division 3	0.57	0.60	0.00	0.00	0.44	0.67	0.36	0.42	0.00	0.71	0.66
Division 4	0.65	0.81	0.67	0.00	0.18	0.59	0.51	0.47	0.47	0.67	0.68
Division 5	0.52	0.90	0.83	0.00	0.29	0.53	0.48	0.18	0.58	0.73	0.71
Division 6	0.66	1.00	1.00	0.17	0.32	0.61	0.58	0.52	0.48	0.71	0.71
Division 7	0.44	0.77	1.00	0.18	0.36	0.70	0.51	0.41	0.18	0.78	0.78
Division 8	0.58	0.00	1.00	0.51	0.00	0.78	0.51	0.58	0.18	0.74	0.74
Division 9	0.53	0.78	-	0.63	0.56	0.77	0.59	0.30	0.18	0.63	0.71
Division 10	0.62	0.81	0.56	0.66	0.68	0.69	0.25	0.00	0.86	0.71	0.70
Division 11	0.54	0.91	0.67	0.69	0.42	0.59	0.51	0.22	0.00	0.73	0.56

The DAPC analysis shows the 136 daffodil cultivars group within 3 clusters (Figure 5-3). Cluster 2 corresponds to daffodil cultivars with parentage belonging to section *Bulbocodii* and cluster 3 to daffodil cultivars with parentage from section *Meridionalis,* corroborating with groups identified in the *matK* haplotype network. Cluster 1 corresponds to all the remaining daffodil cultivars sampled in this study.

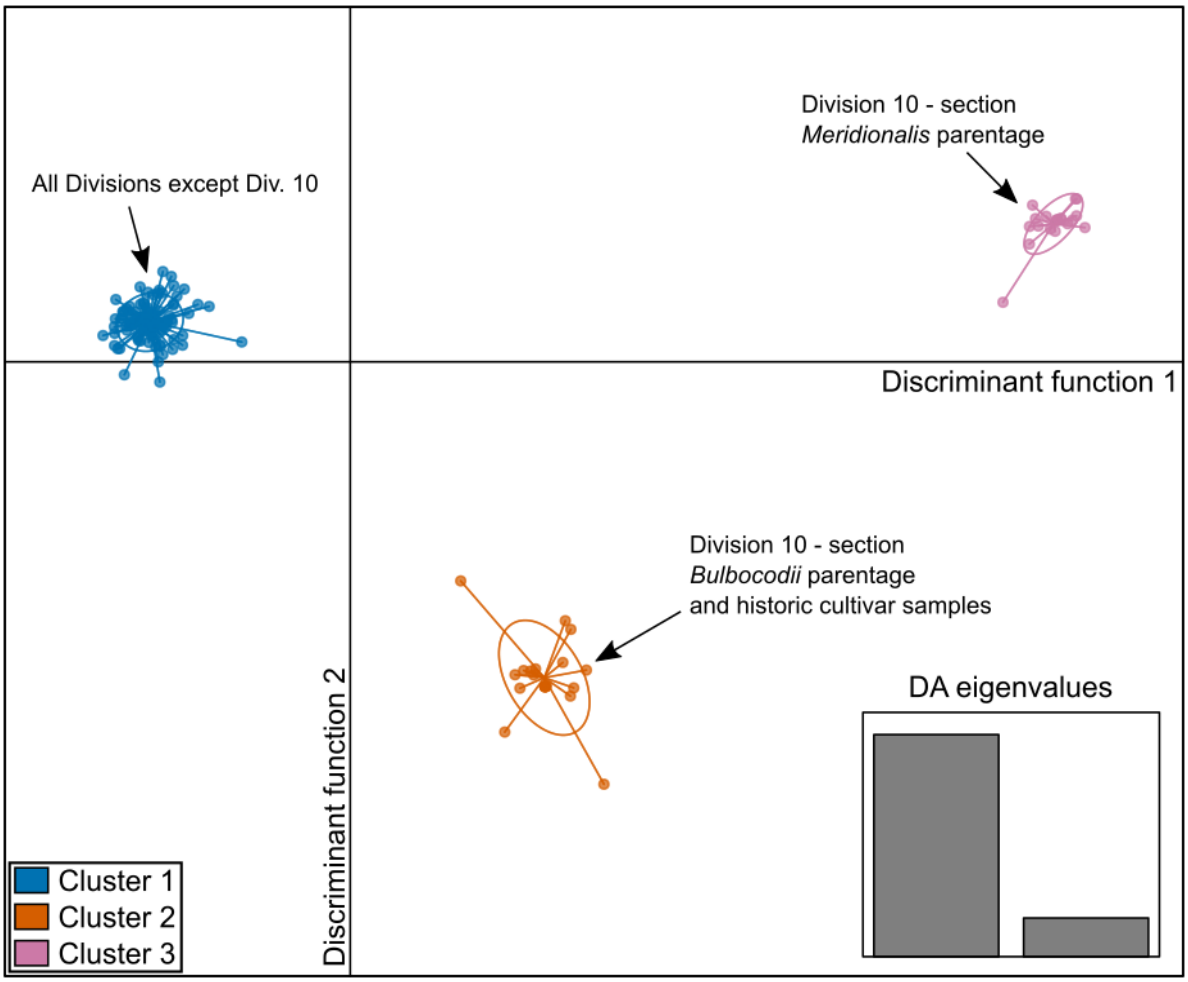


Figure 5-3 Scatterplot of DAPC analysis for 136 genotyped daffodil cultivars when K=3. All principal components were retained. The cluster assignment is plotted on Figure 5-4.

The Neighbor Joining tree shows that only Division 10 was fully recovered (Figure 5-4), whereas the remaining horticultural Divisions do not form clusters based upon genetic distance. The clusters identified in the DAPC analysis have been overlaid onto the NJ tree to identify groupings. Nine cultivars from horticultural Divisions 1, 3, 4, 6, and 9 are nested within cluster 2 in the NJ tree, although these cultivars cluster with Bulbocodium cultivars group 2 of the DAPC analysis. All of these cultivars were sourced from Croft 16 a national collection of heritage daffodil varieties. The one sample of *N.* 'Atlas Gold' is clustered amongst *N.* 'Julia Jane' cultivars in Figure 5-4.

There is zero genetic distance between 10 couplets of cultivars (Figure 5-4): *N.* 'Strines' and *N.* 'Orange Tint', *N.* 'Wheal Bush' and *N.* 'Canary', *N.* 'Pollys Pearl' and *N.* 'Kabani', *N.* 'Papillon Blanc' and *N.* 'Phantom', *N. poeticus* 'Flore Pleno' and *N.* 'Horace', *N.* 'Mission Bells and

N. ‘Puppet’, *N.* ‘Yellow Cheerfulness’ and *N.* ‘Dallas’, *N.* ‘Grand Primo Citroniere’ and *N.* ‘Hexameter’, *N.* ‘Spoirot’-JP and *N.* ‘Spoirot’-JB, *N.* ‘Julia Jane’-MB and *N.* ‘Julia Jane’-BB. The cultivars groups that were found to have no genetic distance account for 12% of sampled cultivars when removing the same cultivar from different suppliers.

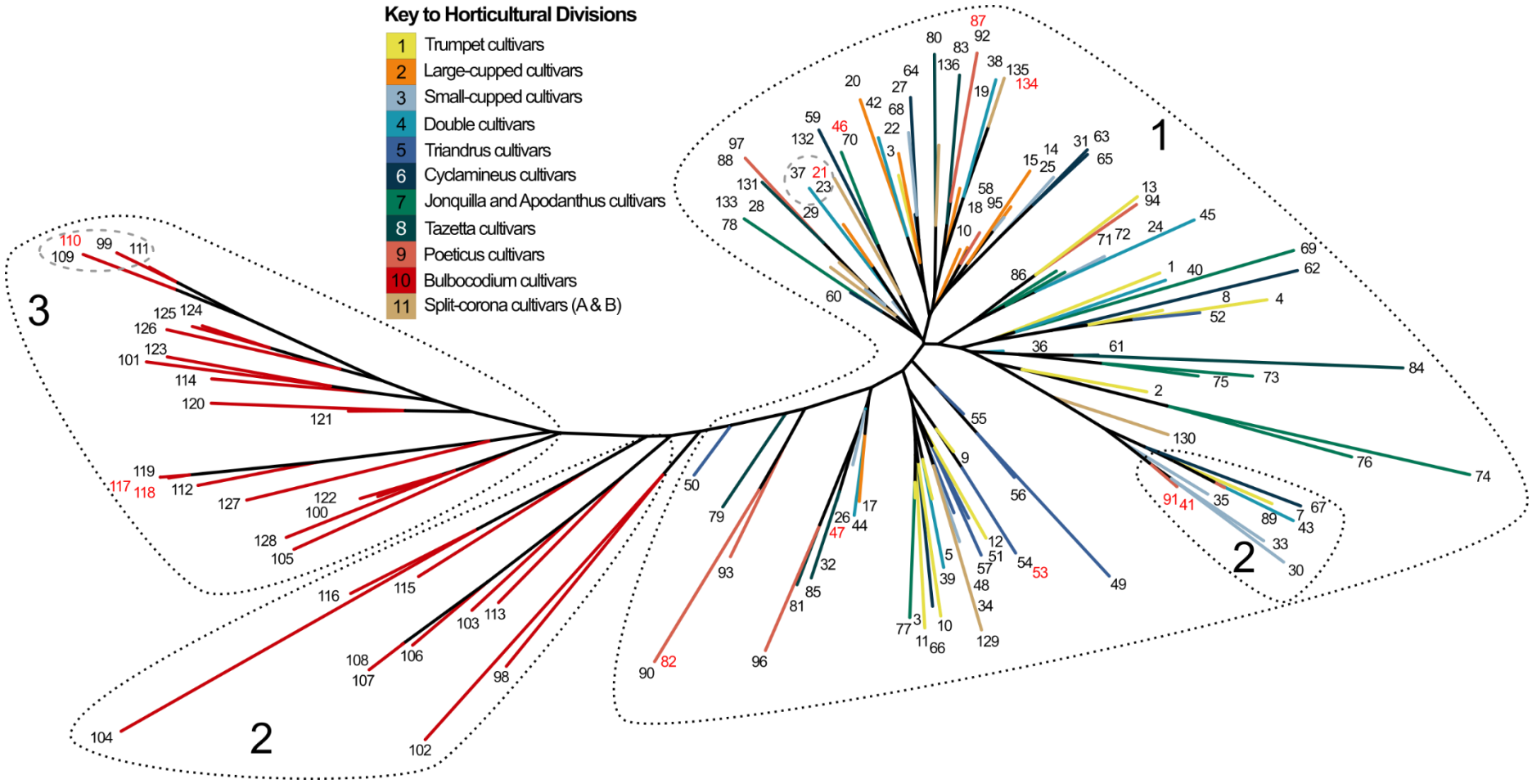


Figure 5-4 Neighbor Joining tree of sampled daffodil cultivars based on Provesti's distance matrix. Line colours correspond to horticultural Division. Cultivar numbers correspond to Table 5-1, numbers in red have zero distance from nearest cultivar. Cultivars with solid black circles correspond to K = 3 cluster scenario. Cultivars circled in grey are discussed in text.

5.5 Discussion

Despite the popularity and utility of microsatellites for analysing population level variation, cross-amplification to even closely related species can be problematic (Bravo, Hoshino, et al., 2006). However, the cross-amplification of microsatellites designed from hoop-petticoat daffodils to all horticultural Divisions was effective with nine of the markers amplifying in more than 90% of sampled daffodil cultivars. The utility of the markers for other daffodil species has already been shown, with successful cross amplification ranging from 39-100% in congeneric *Narcissus* species (Könyves, David, et al., 2016). The transferability of additional daffodil microsatellite loci has been tested on congeneric species, with transferability ranging from 38-100% (Simón, Xavier Picó, et al., 2010), and 0-100% with a small sampling range including daffodil cultivars (Hodgins, Stehlik, et al., 2007). Three additional groups of microsatellite loci have not been used for daffodil cultivar identification to date (Hodgins, Stehlik, et al., 2007, Simón, Xavier Picó, et al., 2010, Barranco, Simón-Porcar, et al., 2019). The cross-amplification of 6 microsatellite loci designed for *N. triandrus* have been shown to successfully cross-amplify in two daffodil cultivars (Hodgins, Stehlik, et al., 2007).

Using the microsatellite data, Division 10 was the only horticultural Division that could be fully recovered. This is not surprising because some Divisions are not derived from a single species or section but are based upon flower morphology (Kington, 2008). An example is the group circled in Figure 5-4 which includes three cultivars from two horticultural Divisions: *N.* 'Strines' (Division 2), *N.* 'Orange Tint' (Division 2), and *N.* 'Angkor' (Division 4), shown in Figure 5-5. Divisions 1-3 are based upon measurements of the corona in comparison with the perianth, and Division 4 contains cultivars with a double perianth, and/or corona. They therefore can be descendants of any *Narcissus* species, providing they have the required floral traits. The pedigree of the three cultivars highlighted represent the complexity of more than 150 years of daffodil breeding programmes (Könyves, David, et al., 2016), with *N.* 'Maximus' (Division 1), *N.* 'Emperor' (Division 1), *N.* 'Kings Court' (Division 1) and *N.* 'King Alfred' (Division 1) repeatedly used within 6 generations of the three cultivars (Daffseek, 2018). The complexity is compounded by the contribution subsection *Pseudonarcissi* has made to daffodil cultivar breeding (Rivera Nuñez, Obón De Castro, et al., 2003).

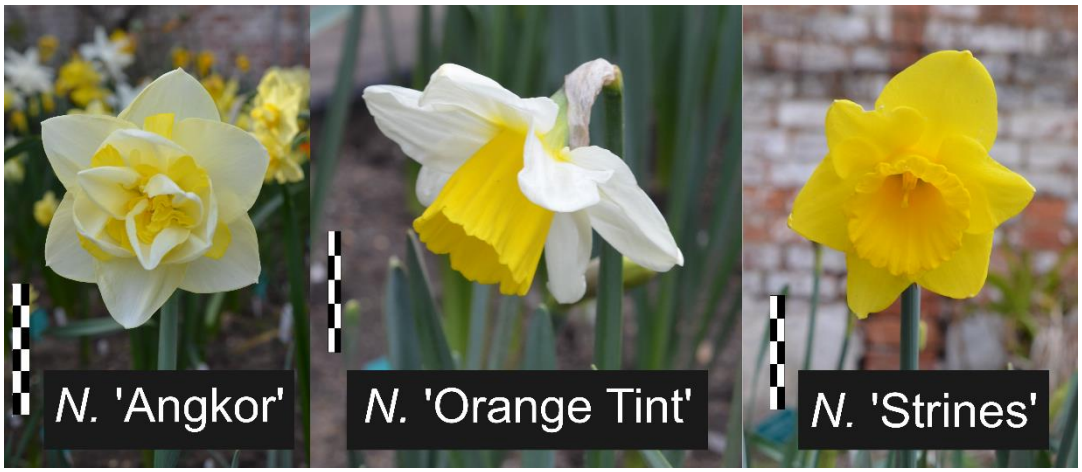


Figure 5-5 Images of daffodil cultivars: *N.* 'Angkor' (Division 2), *N.* 'Orange Tint', and *N.* 'Strines' (both Division 2). Each unit of the scale bar equals 1cm.

The evidence of the contribution of subsection *Pseudonarcissi* has had to daffodil cultivar breeding is highlighted in the *matK* haplotype network, with 64% of cultivars sampled here clustering in H1, corresponding to section *Narcissus*, which includes subsection *Pseudonarcissi* (Marques, Aguilar, et al., 2017). Representatives from all horticultural Divisions can be found in H1, aside from hoop-petticoat daffodil cultivars. This would imply that the hybridisation of hoop-petticoat daffodils with species in other *Narcissus* sections is rare, or that they have solely been used as the pollen parent. The separation of the hoop-petticoat daffodil cultivars into two haplotype groups (H4-7 and H10) corresponds to the splitting of section *Bulbocodii* into two sections: *Meridionalis* and *Bulbocodii* (Marques, Aguilar, et al., 2017). This is supported by the DAPC analysis (Figure 5-3; Figure 5-4) were section *Meridionalis* corresponds to cultivars within cluster 3 and section *Bulbocodii* to cultivars within cluster 2. All the daffodils with *N. bulbocodium* parentage are within cluster 2 of the DAPC plot, with the exception of cultivars with *N. bulbocodium* as a seed parent and *N. cantabricus* as the pollen parent (*N.* 'Kholmes and *N.* 'Spoirot'). Division 8, the Tazetta daffodil cultivars, includes descendants from *Narcissus* section *Tazettae*, within subgenus *Hermione* (Santos-Gally, Vargas, et al., 2012), and therefore the distance of the cultivars in this Division from other daffodil cultivars is unsurprising. However, there are *Tazetta* daffodil cultivars derived from crosses of *N. tazetta* with *N. poeticus*, or *N. jonquilla*. The *Tazetta* cultivars cluster into a distinct group, with one

exception found in H1, *N.* 'Saint Agnes' is the descendant of *N. tazetta* and *N.* 'Chaucer' (Daffseek, 2018), a Division 9 cultivar, which explains clustering with remaining Division 9 daffodil cultivars. The remaining Tazetta Division cultivars were separated in the haplotype network by their parentage with *N. elegans* corresponding to H12 and *N. papyraceus* to H13.

Narcissus 'Cedric Morris' (95), is a selection of *Narcissus minor* L. (subsect. *Pseudonarcissi*), reported to originally have been collected from Costa Verde in Asturias (Blanchard, 1990). The *matK* haplotype H9 (Figure 5-2) was found in *N.* 'Cedric Morris', *N.* 'Lily May', and *N.* 'Rip Van Winkle'. This haplotype corresponds to subsection *Pseudonarcissi* taxa from Galicia and Northern Portugal (Chapter 2), highlighting the provenance of *N.* 'Cedric Morris', *N.* 'Lily May' and *N.* 'Rip Van Winkle'.

The hoop-petticoat cultivar data was gathered in Könyves (2014), and has subsequently been compiled with the data from all other horticultural Divisions here. This factor could explain the dissimilarity, as interpretation errors of up to two repeat lengths have been reported from 6 laboratories analysing identical microsatellite data (de Valk, Meis, et al., 2009). However, this was corrected using standardized allelic sizing ladders and resolution of interpretative problems. In order to overcome these issues the same allelic sizing ladder (Liz 500, Thermo-Fisher Scientific, Paisley, UK), and scoring protocol (Pfeiffer, Roschanski, et al., 2011) was used in this study. Raw data from Könyves (2014) was reanalysed considering the whole data set when scoring peaks, to overcome differences in scoring.

Using the microsatellite data only 12% of the daffodil cultivars here had identical genotypes, although how consistent the genotypes are scored for multiple accessions of the same cultivar was limited to only four cultivars in this study. Multiple samples sourced of the same cultivar, *N.* 'Julia Jane', *N.* 'Tarlatan', *N.* 'Spoirot', and *N.* 'Golden Bells' all clustered closely. To maintain cultivar uniformity clonal propagation is advised, however it is known that cultivars which produce few bulb offsets have previously been propagated through seed (Hanks, 2002b, Kington, 2008). Issues could therefore arise identifying daffodils as the ICNCP (Brickell, Alexander, et al., 2016) states that cultivars should be 'distinct, uniform, and stable'. There is one discrepancy to the clustering of the same cultivars from multiple sources, as *N.* 'Apollo

Gold' can be found nested amongst *N.* 'Julia Jane' accessions, both cultivars are selections of *N. romieuxii* (Daffseek, 2018).

Narcissus 'Atlas Gold' was grouped among the cultivars of *N.* 'Julia Jane', a selection of *N. romieuxii* in the NJ analysis (Figure 5-4), and its multilocus genotype was most similar to the *N.* 'Julia Jane' accession from Pottertons Nursery (PM). Upon closer inspection, it is evident that *N.* 'Julia Jane' (PM) is mislabelled stock of *N.* 'Atlas Gold', which was obtained from the same source (Könyves, 2014). Plants of *N.* 'Julia Jane' (BB and MB) have a light-yellow flower, while *N.* 'Julia Jane' (PM) and *N.* 'Atlas Gold' have a bright yellow flower. According to Kington (2008) these two cultivars differ in their flower colour, with the RHS colour chart values of 2C or 3C for *N.* 'Julia Jane', and 6A for *N.* 'Atlas Gold'. Moreover, both of these are selections of *N. romieuxii* collected from the wild by J.C. Archibald, and from the same collection as evident by the seedling numbers JCA805Y (*N.* 'Atlas Gold') and JCA805 (*N.* 'Julia Jane') (Kington, 2008, Daffseek, 2018). This shows that the microsatellite markers developed here have the potential to be used to identify mislabelled bulb stocks, even without observing the flower colour.

Fingerprinting methods have been used successfully for cultivar identification in groups such as carnations (Arens, Esselink, et al., 2009), chrysanthemums (Kjos, Fjellheim, et al., 2010) , *Lagerstroemia* (Cai, Pan, et al., 2011), and to observe pedigree structure in crops such as peaches (Sitther, Road, et al., 2012) and apples (Moriya, Iwanami, et al., 2011). However, the pedigree of daffodil cultivars is often much more complex (sample pedigree chart - Appendix B), and there are fewer currently available genetic resources than are available for model organisms or crops such as apples. The markers designed for hoop-petticoat daffodils are shown here to be appropriate for hoop-petticoat daffodil cultivar identification. There is clustering of daffodil cultivars from all horticultural Divisions based upon their pedigree, indicative of the utility of microsatellites for daffodil cultivar identification. Moreover, the microsatellite loci can successfully identify mislabelled stocks of daffodil cultivars, demonstrating their utility for identifying mis-sold cultivars and help in identifying synonymy amongst the more than 30,000 registered cultivars. However, a suite of microsatellite loci will need to be developed to provide the resolution to unpick the complex breeding patterns.

Evidence here suggests microsatellite loci would need to be developed utilising the whole of *Narcissus*, with each section represented, in order to create a robust marker set.

Next-Generation sequencing has been utilised to produce plastid genomes for representatives across the genus, allowing a marker set to be mined capable of cross amplifying across all cultivars and possessing the required levels of variation (Chapter 4). The application of NGS is now becoming a reality in other ornamental crops such as *Dianthus* (Yagi, Kosugi, et al., 2014), *Petunia* (Bombarely, Moser, et al., 2016), and *Hibiscus* (Kim, Kim, et al., 2017). Alternatively, with the rapid uptake of genome sequencing, thousands of single nucleotide polymorphisms (SNPs) could be identified. SNPs are widely abundant across the genome, and allow for automated analysis of samples (Cabezas, Ibáñez, et al., 2011). This would be beneficial should a molecular means for daffodil cultivar identification be utilised in industry. Recently *Narcissus poeticus* became the first published daffodil plastid genome (Könyves, Bilsborrow, et al., 2018), opening the opportunity to utilise genome and partial genome datasets for daffodil cultivar identification in the near future.

Chapter 6 General Discussion

The principal aim of this thesis was to undertake a detailed study of the species boundaries in the *N. minor* group and closely related species in subsection *Pseudonarcissi*, using both DNA sequencing and microsatellite analysis. Subsection *Pseudonarcissi* is the most economically important group of species, since trumpet daffodils have made one of the largest contributions to the number of daffodil cultivars. This project used the *N. minor* group as an example to explore species circumscription in the trumpet daffodils as it is relatively well-defined and includes a number of horticulturally important taxa. This lead to similar methods being applied to observe to what extent daffodil cultivars could be identified.

In line with botanical study, morphological data have been most widely used in the past to determine species boundaries for taxonomic investigations of *Narcissus* (Fernandes, 1968, Webb, 1980, Blanchard, 1990, Aedo, 2013). Discontinuous characters, such as number of flowers per scape, have been used to identify well-defined sections or species groups, but analysis of continuous morphological characters is often applied for species description within groups. This is an artefact of a limited number of characters in use for identification. An example of this is the identification of the *N. minor* group, where quantitative characters are used to identify species (Blanchard, 1990). In *Narcissus,* this has often led to arbitrary divisions in continuous variation being used for the description of species, which is not reflected in molecular variation. Species have been poorly defined using characters that are part of a continuous morphological variation, therefore 'splitting' has consequently led to the wide-range in taxa that have been accepted in *Narcissus*: 26 (Webb, 1980) to 86 (Mathew, 2002).

Morphology-based classification at section level provides well-supported evidence for ten sections (Fernandes, 1968, Webb, 1980, Blanchard, 1990). Cytological evidence has been used to define 12 sections of *Narcissus* (Zonneveld, 2008), splitting *Juncifolii* from *Jonquillae* and *Nevadensis* from *Pseudonarcissi*. The use of phylogenetic analysis has confirmed many of the sections that are based upon cytological results. However, several sections are not identified as monophyletic (Marques, Aguilar, et al., 2017), such as sections *Pseudonarcissi* and *Bulbocodii* despite being considered morphologically distinct sections. One clade within subsection *Pseudonarcissi* corresponds to the *N. minor* group, *N. cyclamineus* and several other taxa within subsection *Pseudonarcissi*, and this thesis reveals strong support for the group based

upon molecular grounds. However, identification of the species included within the *N. minor* group was not possible based on molecular evidence, and *N. cyclamineus* not previously thought to belong in the group is found to be nested within the clade.

This pattern could be explained by the prevalent hybridisation that has been reported in *Narcissus*, even occurring between species of different subgenera (Marques, Feliner, et al., 2010), widespread hybridisation across the genus is evident by the number of reported hybrids (Fernández Casas, 2013). Hybridisation events often lead to offspring with intermediate morphological characters (Yakimowski and Rieseberg, 2014), which can confuse species boundaries when applying the morphological species concept. Recently diverged species provide difficulty for the determination of species boundaries as populations may not be reproductively isolated, leading to a lack of molecular differentiation (Shaffer and Thomson, 2007). To overcome this issue, population genetic approaches are appropriate, and have been used previously to ascertain species boundaries in hoop-petticoat daffodils (Könyves, 2014), trumpet daffodils (Jiménez, Sánchez-Gómez, et al., 2009), or intraspecific variation in *N. triandrus* and *N. papyraceus*. (Hodgins and Barrett, 2007, Simon-Porcar, Pico, et al., 2015). However, such approaches did not resolve species of the *N. minor* group based upon molecular grounds in this project.

Hybridisation of *Narcissus* species by plant breeders has given rise to thousands of cultivars, that highlight the lack of sexual barriers and level of gene exchange possible amongst daffodils (Brandham, 1987). One of the barriers for hybridisation in *Narcissus* is therefore the seasonality of species, although spring flowering is common; there are several autumn flowering species, such as *N. cavanillesii*, *N. serotinus*, *N. viridiflorus* (Blanchard, 1990). However, there have been artificial crosses of *N. viridiflorus* and spring flowering species in cultivation, one example being *N.* 'Lima's Green Road' (*N. viridiflorus* × *N.* 'El Camino'). In the wild flowering time would provide a breeding barrier limiting hybridisation. However, there are groups of daffodils found flowering outside 'typical' flowering season if the conditions are right, such as hoop-petticoat daffodils (Könyves, 2014). The variation in flowering time means hybridisation can occur between taxa that would not usually cross in *Narcissus* due to seasonality.

Beside seasonality, pollinators of *Narcissus* provide another breeding barrier. There is variation in the breeding systems of daffodils with some heterostylous species (e.g. section *Ganymedes*) and mechanisms for self-incompatibility (e.g. section *Bulbocodii*). A range of pollinators has been reported in daffodils: solitary bees, diurnal moths, hoverflies, and pierid butterflies (Pérez-Barrales, Vargas, et al., 2004). Shifts from style dimorphism have been shown to correlate with shifts to bee pollination in *Narcissus* sect. *Apodanthi,* whilst moth pollination sustained dimorphism (Pérez-Barrales, Vargas, et al., 2006). This adaptive evolution for floral shape suggest selective pressure on floral characters based upon pollinators in *Narcissus*. Widely distributed species often have varying pollinators across their range, populations of *N. papyraceus* were found to have varying morphological traits depending upon the pollinator (Pérez-Barrales, Arroyo, et al., 2007). This provides evidence that pollinators provide selection for stabilised floral morphological variation forms, which could have given rise to the recognition of species through arbitrary divisions in continuous variation. Although the pollinators of the *N. minor* group are not known, a pollinator study of *N. cyclamineus* reported several bee species and hoverflies visit the plants (Navarro et al 2012).

The most likely factor providing the barriers to gene flow is geography; the present study shows that genetic distance correlates with geographic distance within the *N. minor* group. This explains the geographic structure observed in molecular studies of *Narcissus*. Molecular variation patterns identified in previous studies of *Narcissus* conflict with the morphological data for certain species; using both phylogenetic (Graham and Barrett, 2004, Marques, Aguilar, et al., 2017, Könyves, David, et al., 2019) and population level approaches (Jiménez, Sánchez-Gómez, et al., 2009, Medrano, López-Perea, et al., 2014). Both seed dispersal and movement of pollen have been shown to be localised in *Narcissus* species such as: *N. longispathus* and *N. assoanus* (Baker, Barrett, et al., 2000, Barrett, Cole, et al., 2004), localised gene flow correlate with the geographic structure observed. Such geographic structure could be expected within panmictic populations of a species; in *Ipomoea pres-caprae* seed dispersal was found to maintain a geographic population structure of three distinct groups (Miryeganeh, Takayama, et al., 2014). However, in daffodils there is a geographic structure to populations evident over several species. This is not unique to *Narcissus*, a geographic molecular structure has been

found in *Allium* section *Sikkimensia* (Xie, Xie, et al., 2019), where haplotypes correspond to a specific mountain range rather than morphological based species. The geographic molecular patterns observed in *Narcissus* could be explained by the gene flow amongst several species within geographic regions that prevents molecular patterns corroborating morphological based species.

Some species remain morphologically distinct entities, despite the reported gene flow. This is evident in the hoop-petticoat daffodils, where section *Bulbocodii* corresponds to a morphological distinct group, easily distinguished by a funnel-shaped corona and reduced tepals. However, molecular patterns provide evidence for the section to be divided in section *Meridionalis* and *Bulbocodii*, that correlate strongly with the distribution of species in northern Africa and the Iberian Peninsula (Marques, Aguilar, et al., 2017, Könyves, David, et al., 2019).

Aside from hybridisation the molecular patterns observed in *Narcissus* are likely to be due to the recent diversification of species within the genus. The only published dated phylogeny, provides evidence that the majority of sections and species groups within the genus have diversified 5-6 Mya (Santos-Gally, Vargas, et al., 2012). This indicates rapid morphological diversification of the genus, which cannot be identified on molecular grounds due to incomplete lineage sorting. Rapid diversification has been found in *Lupinus* species throughout the Andes, giving rise to wide morphological variation that is not reflected in molecular data, it is proposed that the major driver in *Lupinus* diversification is island-like habitats (Hughes and Eastwood, 2006). Similarly, molecular variation did not corroborate wide morphological diversification in *Aquilegia* with differing mechanisms for diversification of European to North American species, diversification of European species was due to isolation and habitat shifts, whilst in north America the diversification was due to pollinator shifts (Bastida, Alcántara, et al., 2010). These studies highlight the complex life histories of plants that have undergone rapid diversification, which has led to morphological diversification without reflecting species within lineages.

Surprisingly, *N. cyclamineus* could not be characterised using molecular microsatellite data, despite being easily distinguished morphologically in subsection *Pseudonarcissi*, where morphological separation is challenging (Blanchard, 1990). Phylogenetic study of plastid gene

regions has also failed to separate *N. cyclamineus* from the *N. minor* group (Graham and Barrett, 2004, Marques, Aguilar, et al., 2017, Könyves, David, et al., 2019). Unfortunately, only one sample of *N. cyclamineus* was included for the only published nuclear gene tree (Marques, Aguilar, et al., 2017), and therefore there are limitations to the conclusions drawn from published data on the monophyly of the species. Similar patterns of discordance between morphology and molecular data have been identified within the *Lotus corniculatus* complex, in the case of morphologically distinct species (Kramina, Meschersky, et al., 2018), a combination of both data-sets were therefore used for species circumscription. Morphology and molecular approaches having conflicting results is not limited to flowering plants, with these issues identified in Bryophytes (Sotiaux, Enroth, et al., 2009) and Ferns (Dubuisson, Hébant-Mauri, et al., 1998), and even more distantly related with examples in marine organisms (Debiasse and Hellberg, 2015), and birds (Mayr, 2011).

Measurements and floral characters have been used to classify species in *Narcissus* in the past (Fernandes, 1968, Webb, 1980, Aedo, 2013), with a heavily reliance upon continuous variation. There have been no in-depth morphometric approaches for species recognition to date, although this work was not the intended for this thesis. Using a combination of linear morphometric analysis and more sophisticated geometric morphometric approaches would be useful to identify morphological boundaries and overlap of variation in *Narcissus.* Similar approaches have been taken in population studies of two Araceae species using leaf landmarks (Souza Silva, Moreira de Andrade, et al., 2012), although daffodil leaves are more simple, therefore a combination of floral, leaf and other characters such as leaf cross-sections could be used. These approaches have demonstrated using toothed leaf margin area for *Tilia* identification (Corney, Tang, et al., 2012) and leaf venation characters for *Passiflora* identification (De Oliveira Plotze and Bruno, 2009).

With the depreciating costs of Next-Generation sequencing, larger datasets are more accessible for species circumscription and marker development. Importantly regions from multiple plant genomes can be acquired through genome skimming, to identify gene flow arising from pollen and seed dispersal (Dodsworth, 2015). The ability to mine genes from high copy regions of multiple genomes makes the method applicable to untangle the prevalent

hybridisation within *Narcissus*. The first plastome for *Narcissus* was published from data gathered in this thesis, the first for subfamily Amaryllidoideae. This has provided a robust reference plastome to assemble the genomes of other daffodil species sequenced for this thesis and can be used for further sampling. The ten daffodil plastomes provide a rich genetic resource for mining variable regions to develop robust markers for both species delimitation and the development of molecular tools for daffodil cultivar identification.

Phylogenetic analysis in *Narcissus* has previously used only a handful of genes to elucidate species boundaries (Graham and Barrett, 2004, Santos-Gally, Vargas, et al., 2012, Marques, Aguilar, et al., 2017, Könyves, David, et al., 2019). The most comprehensive phylogeny to date is based upon sequence data from all three plant genomes, however this equates to approximately 4,000 bp of sequenced DNA. The plastomes developed in this project provide the initial step in developing 'super barcodes' (Li, Yang, et al., 2015) for species circumscription. The phylogenies constructed from whole plastomes and mined nuclear ribosomal regions reveal the placement of section *Apodanthi* is unresolved. Mitochondrial reads have been successfully used to resolve the placement of species in Oleaceae when plastomes were not able to do so (Van de Paer, Bouchez, et al., 2017). Given mitochondrial genes can be mined from genome skim data (Ripma, Simpson, et al., 2014), this could be an initial step to resolve the placement of section *Apodanthi*. This could be completed from the raw data collected for this thesis, however there was not enough time to construct the mitochondrial scaffolds during this project. Conflicting patterns of tree topology based upon different genome sequences have been identified in *Narcissus* (Marques, Aguilar, et al., 2017), conflict of plastid and nuclear genomes have been shown to give rise to conflicting topologies in *Ipomoea*, with chloroplast capture the likely explanation (Muñoz-Rodríguez, Carruthers, et al., 2018).

The sampling was chosen in the plastid genomics study to encompass the widest possible range of variation within *Narcissus*, identified from published phylogenies (Könyves, 2014, Marques, Aguilar, et al., 2017) and data in this thesis. Should further microsatellite loci be developed, the wide sampling of the genus means common pitfalls with successful cross-amplification of variable markers such as microsatellites should be limited. However, increasing

the sampling should be a key next step of the overall project to identify sectional placement and monophyly of sections.

Subsequently, sequencing transcriptomes in *Narcissus* would be worthwhile to develop markers from expressed genes. There are issues of low genetic variation in widely used plastid genes, for complex species groups like *Salix* (Percy, Argus, et al., 2014), and *Calligonum* (Li, Feng, et al., 2014), due to the propensity of hybridisation exhibited in plants. Building transcriptomes could provide another genetic resource for marker development, transcriptome based SSRs have been used to identify species of *Argyranthemum* (White, Doo, et al., 2016), when universal markers were hampered by a lack of variation. Similar approaches could be applied in *Narcissus*, to observe the correlation between genetic variation in functional genes and observed morphology. This approach of transcriptome analysis has been demonstrated in *Petunia*, revealing genes involved in morphological differentiation of flowers (Amano, Kitajima, et al., 2018), which could be used to explain why morphological variation in *Narcissus* is often not reflected in molecular variation. Transcriptomics would therefore be a practical next step for *Narcissus*, however whole genome sequencing is becoming increasingly widespread.

There are ambitious large-scale NGS projects already underway such as, PhyloAlps (Coissac, Boyer, et al., 2018), and 10KP (Cheng, Melkonian, et al., 2018). After surpassing 1,000 plant transcriptomes (Matasci, Hung, et al., 2014), the project 10KP has been developed to assemble whole genomes of samples from every major clade of plants by 2022. This large scale genome data can be used to explain mechanisms for plant evolution. Although, 10KP currently focuses upon plants with small diploid genomes (usually less than 2C=5 Gbp), with daffodil genomes being much larger (2C=13-66 Gbp). As our knowledge of genome wide data develops, the techniques will be applied more widely and to more complex genomes. Pipelines are already in development for the assembly of the *Tulipa gesneriana* genome (2C=34 Gbp) from long range sequencing, although this is currently both complex and expensive. The trumpet daffodils have made a large contribution to daffodils in horticulture (Rivera Nuñez, Obón De Castro, et al., 2003). Understanding the levels of variation within subsection *Pseudonarcissi* is therefore key to identifying variation within cultivation. Additional aims in this research were to identify to what extent currently available microsatellites could identify daffodil cultivars.

Microsatellites have been successfully used for cultivar identification in other ornamental plants such as *Anthurium* (Wang and Chuang, 2013), *Dianthus* (Yagi, Kosugi, et al., 2014) and *Pelargonium* (Peltier, Becher, et al., 2000). A reliable method for cultivar identification is required to keep names consistent (Culham and Grant, 1999), and provide gardeners with the characteristics they expect from their chosen plant varieties. Given the number of daffodil cultivars registered, a molecular means for identification is required as current morphological approaches are not capable of identifying the distinguishing characteristics for all cultivars.

Published microsatellites designed for hoop-petticoat daffodils were successfully applied to daffodils from all horticultural Divisions, to establish the effectiveness for cultivar identification. Over 80% of the cultivars screened had a unique genetic signature. Hoop-petticoat daffodil cultivars were divided into two groups based upon parentage, this coincides with the recent separation of wild species in section *Bulbocodii* to sections: *Meridionalis* and *Bulbocodii* (Marques, Aguilar, et al., 2017). Results from the remaining daffodil cultivars reflect the complex breeding history which has taken place over the past 150 years in particular (Brandham, 1987). Although it was not expected to resolve cultivars at the horticultural Division level, as some Divisions are based upon artificial criteria (e.g. double daffodil cultivars) which bear no direct relationship to evolutionary origin. However, several Divisions that are based upon characteristics of species or sections were not recovered; this could be due to the lack of resolution at species level although it could also be caused by the prevalent hybridisation between species of different sections. Consequently, the results from this thesis indicate that a suite of microsatellites would need to be developed considering the whole genus. There are already published additional microsatellites (Hodgins, Stehlik, et al., 2007, Simón, Xavier Picó, et al., 2010), developed using *N. triandrus* (section *Ganymedes*), and *N. papyraceus* (section *Tazettae*), which could be applied for daffodil cultivar identification. The additional microsatellites could potentially identify cultivars from Division 5 (*Triandrus* daffodil cultivars) and Division 8 (*Tazetta* daffodil cultivars).

6.1 Conclusions

This project has demonstrated the utility of microsatellites for cultivar identification in a complex horticultural group. Furthermore, this thesis highlights the challenges of species circumscription in *Narcissus*, pushing towards fewer recognised species in the genus. The application of microsatellites and DNA sequencing reveal a geographic structure that does not support morphological based species in the *Narcissus minor* group or *N. cyclamineus*. To identify regions of genomic variability the first whole plastomes for *Narcissus* have been assembled. A comparative analysis has been reported, providing a rich genetic resource that can be used for marker development for future circumscription of species and for the development of molecular tools for cultivar identification. Harnessing larger datasets acquired through high-throughput sequencing will be key to untangle the complex molecular patterns for future *Narcissus* research. The research reported in this study provides the first step in developing molecular tools for species circumscription and authenticating daffodil cultivars in horticulture.

The depreciating costs of Next-Generation sequencing over the past decade and rapid development of computer power and the software to handle large data sets has led to the access of genome wide data. This will continue to become commonplace in molecular studies, even though whole genome sequencing is complex, for large genomes, and still currently expensive. As our understanding of whole genome data continues, long-read sequencing develops, and with the rapid uptake of the techniques, whole genome data are likely to be widely accessible in the next decade.

Molecular data approaches had previously been thought to give unequivocal resolution of species relationships, this work challenges that assumption. Future work should therefore investigate the reasons behind the observed patterns of molecular variation.

References

A.P.G III, 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), pp. 105–121.

Aedo, C., 2010. Typyfications of the names of Iberian accepted species of *Narcissus* L. (Amaryllidaceae). *Acta Botanica Malacitana*. 35 (1980), pp. 113–122.

Aedo, C., 2013. *Narcissus* L. In: S. Talavera, C. Andrés, M. Arista, M.P.F. Piedra, E. Rico, M.B. Crespo, A. Quintanar, A. Herrero and C. Aedo, eds. Flora iberica. Plantas vasculares de la Península Ibérica e Islas Baleares. Vol. XX, Liliaceae-Agavaceae. Madrid: Real Jardín Botánico.

Amano, I., Kitajima, S., Suzuki, H., Koeduka, T., and Shitan, N., 2018. Transcriptome analysis of *Petunia axillaris* flowers reveals genes involved in morphological differentiation and metabolite transport. *PLOS ONE*. 13 (6): e0198936.

Amiryousefi A., Hyvönen J., Poczai P., 2018. Genome analysis IRscope: an online program to visualize the junction sites of chloroplast genomes. *Bioinformatics*. Advance Access.

Arens, P., Esselink, D., Noordijk, Y., Kodde, L., Hof, L., Wietsma, W., and Vosman, B., 2009. Identification of Carnation Varieties Using Microsatellite Markers. *Acta Horticulturae*. (836), pp. 175–181.

Arroyo, J. and Barrett, S.C.H., 2000. Discovery of distyly in *Narcissus* (Amaryllidaceae). *American Journal of Botany*. 87 (5), pp. 748–751.

Avis-Riordan, K., 2017. This if officially Britain's most popular flower. Country Living [online]. Available from: https://www.countryliving.com/uk/homes-interiors/gardens/news/a1838/britains-favourite-flower-vote/ [Accessed 8 Nov 2018].

Ben Ayed, R., Ben Hassen, H., Ennouri, K., Ben Marzoug, R., and Rebai, A., 2016. OGDD (Olive Genetic Diversity Database): A microsatellite markers' genotypes database of worldwide olive trees for cultivar identification and virgin olive oil traceability. Database. 2016, pp. 1–9.

Baker, A.M., Barrett, S.C.H., and Thompson, J.D., 2000. Variation of pollen limitation in the early flowering Mediterranean geophyte *Narcissus assoanus* (Amaryllidaceae). *Oecologia*. 124, pp. 529–535.

Baker, J.G., 1875. Review of the genus *Narcissus*. In: F. Burbridge, ed. The *Narcissus*: its history and culture. London: Reeve. pp. 61–87.

Bakker, F.T., Lei, D., Yu, J., Mohammadin, S., Wei, Z., Van de Kerke, S., Gravendeel, B., Nieuwenhuis, M., Staats, M., Alquezar-Planas, D.E., and Holmer, R., 2016. Herbarium genomics: Plastome sequence assembly from a range of herbarium specimens using an Iterative Organelle Genome Assembly pipeline. *Biological Journal of the Linnean Society*. 117 (1), pp. 33–43.

Baldwin, B.G., Sanderson, M.J., Porter, J.M., Wojciechowski, M.F., Campbell, C.S., and Donoghue, M.J., 1995. The ITS region of nuclear ribosomal DNA: a valuable source of evidence on angiosperm phylogeny. *Annals of the Missouri Botanical Garden*. 82 (2), pp. 247–277.

Barranco, D., Arroyo, J., and Santos-Gally, R., 2019. Avoiding sexual interference: herkogamy and dichogamy in style dimorphic flowers of *Narcissus broussonetii* (Amaryllidaceae). *AoB PLANTS*. 11 (4), pp. 1–10.

Barranco, D., Simón-Porcar, V.I., and Arroyo, J., 2019. Characterization of microsatellite markers for *Narcissus dubius, N. cuatrecasasii, N. assoanus and N. rupicola* (Amaryllidaceae). *Plant Genetic Resources: Characterization and Utilization*. 17 (03), pp. 285–288.

Barrett, S.C.H., Cole, W.W., and Herrera, C.M., 2004. Mating patterns and genetic diversity in the wild Daffodil *Narcissus longispathus* (Amaryllidaceae). *Heredity*. 92 (5), pp. 459–465.

Barrett, S.C.H. and Harder, L.D., 2004. The evolution of polymorphic sexual systems in daffodils (*Narcissus*). *New Phytologist*. 165, pp. 45–53.

Bastida, J.M., Alcántara, J.M., Rey, P.J., Vargas, P., and 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), pp. 171–185.

Beatty, G.E. and Provan, J., 2013. Post-glacial dispersal, rather than in situ glacial survival, best explains the disjunct distribution of the Lusitanian plant species *Daboecia cantabrica* (Ericaceae). *Journal of Biogeography*. 40 (2), pp. 335–344.

Beier, S., Thiel, T., Münch, T., Scholz, U., and Mascher, M., 2017. MISA-web: a web server for microsatellite prediction. *Bioinformatics*. 33 (16), pp. 2583–2585.

Benito Garzón, M., Sánchez de Dios, R., and Sáinz Ollero, H., 2007. Predictive modelling of tree species distributions on the Iberian Peninsula during the Last Glacial Maximum and Mid-Holocene. *Ecography*. 30 (1), pp. 120–134.

Bi, Y., Zhang, M.F., Xue, J., Dong, R., Du, Y.P., and Zhang, X.H., 2018. Chloroplast genomic resources for phylogeny and DNA barcoding: A case study on *Fritillaria*. *Scientific Reports*. 8 (1), pp. 1–12.

Blanchard, J.W., 1990. *Narcissus*: a guide to wild daffodils. Woking: Alpine Garden Society.

Bolger A.M., Lohse M., Usadel B., 2014. Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics*. 30:2114-2120.

Bombarely, A., Moser, M., Amrad, A., Bao, M., Bapaume, L., Barry, C.S., Bliek, M., Boersma, M.R., Borghi, L., Bruggmann, R., Bucher, M., D'Agostino, N., Davies, K., Druege, U., Dudareva, N., Egea-Cortines, M., Delledonne, M., Fernandez-Pozo, N., Franken, P., Grandont, L., Heslop-Harrison, J.S., Hintzsche, J., Johns, M., Koes, R., Lv, X., Lyons, E., Malla, D., Martinoia, E., Mattson, N.S., Morel, P., Mueller, L.A., Muhlemann, J., Nouri, E., Passeri, V., Pezzotti, M., Qi, Q., Reinhardt, D., Rich, M., Richert-Pöggeler, K.R., Robbins, T.P., Schatz, M.C., Schranz, M.E., Schuurink, R.C., Schwarzacher, T., Spelt, K., Tang, H., Urbanus, S.L., Vandenbussche, M., Vijverberg, K., Villarino, G.H., Warner, R.M., Weiss, J., Yue, Z., Zethof, J., Quattrocchio, F., Sims, T.L., and Kuhlemeier, C., 2016. Insight into the evolution of the Solanaceae from the parental genomes of *Petunia hybrida*. *Nature Plants*. 2: 16074, pp. 1–9.

Brandham, P.E., 1987. The Chromosomes of Species, Hybrids and Cultivars of *Narcissus* L. (Amaryllidaceae). *Kew Bulletin*. 42 (1), pp. 65–102.

Bravo, J.P., Hoshino, A.A., Angelici, C.M.L.C.D., Lopes, C.R., and Gimenes, M.A., 2006. Transferability and use of microsatellite markers for the genetic analysis of the germplasm of some *Arachis* section species of the genus *Arachis*. *Genetics and Molecular Biology*. 29 (3), pp. 516–524.

Brickell, C.D., Alexander, C., David, J.C., Hetterscheid, W.L. a., Leslie, A.C., Malecot, V., and Jin, X., 2016. International Code of Nomenclature for Cultivated Plants (Ninth Edition). Scripta Horticulturae. ISHS.

Britton, N.L. and Brown, A., 1913. Illustrated Flora of the Northern United States, Canada, and the British Possessions. 2nd Edition. New York: C. Scribner's sons,1913.

Cabezas, J.A., Ibáñez, J., Lijavetzky, D., Vélez, D., Bravo, G., Rodríguez, V., Carreño, I., Jermakow, A.M., Carreño, J., Ruiz-García, L., Thomas, M.R., and Martinez-Zapater, J.M., 2011. A 48 SNP set for grapevine cultivar identification. *BMC Plant Biology*. 11 (153), pp. 1–12.

Cai, M., Pan, H.T., Wang, X.F., He, D., Wang, X.Y., Wang, X.J., and Zhang, Q.X., 2011. Development of novel microsatellites in *Lagerstroemia indica* and DNA fingerprinting in Chinese *Lagerstroemia* cultivars. *Scientia Horticulturae*. 131 (1), pp. 88–94.

Caldas, F.B. and Moreno Saiz, J.C., 2011. *Narcissus asturiensis*. The IUCN Red List of Threatened Species 2011 [online]. Available from: http://dx.doi.org/10.2305/IUCN.UK.2011-1.RLTS.T162098A5536148.en [Accessed 1 Sep 2018].

Castillo-Mendoza, E., Salinas-Sánchez, D., Valencia-Cuevas, L., Zamilpa, A., and Tovar-Sánchez, E., 2018. Natural hybridisation among *Quercus glabrescens*, *Q. rugosa* and *Q. obtusata* (Fagaceae): Microsatellites and secondary metabolites markers. *Plant Biology*. 21, pp. 110–121.

CBOL Plant Working Group, 2009. A DNA barcode for land plants. *Proceedings of the National Academy of Sciences*. 106 (31), pp. 12794–12797.

Chaïr, H., Sardos, J., Supply, A., Mournet, P., Malapa, R., and Lebot, V., 2016. Plastid phylogenetics of Oceania yams (*Dioscorea* spp., Dioscoreaceae) reveals natural interspecific hybridization of the greater yam (*D. alata*). *Botanical Journal of the Linnean Society*. 180 (3), pp. 319–333.

Chao, Y.-T., Chen, W.-C., Chen, C.-Y., Ho, H.-Y., Yeh, C.-H., Kuo, Y.-T., Su, C.-L., Yen, S.-H., Hsueh, H.-Y., Yeh, J.-H., Hsu, H.-L., Tsai, Y.-H., Kuo, T.-Y., Chang, S.-B., Chen, K.-Y., and Shih, M.-C., 2018. Chromosome-level assembly, genetic and physical mapping of *Phalaenopsis aphrodite* genome provides new insights into species adaptation and resources for orchid breeding. *Plant Biotechnology Journal*. 16 (12), pp. 2027–2041.

Chapman, M.A., 2015. Transcriptome Sequencing and Marker Development for Four Underutilized Legumes. *Applications in Plant Sciences*. 3 (2), p. 1400111.

Chase, M.W., Reveal, J.L., and Fay, M.F., 2009. A subfamilial classification for the expanded asparagalean families Amaryllidaceae, Asparagaceae and Xanthorrhoeaceae. *Botanical Journal of the Linnean Society*. 161 (2), pp. 132–136.

Cheng, S., Melkonian, M., Smith, S.A., Brockington, S., Archibald, J.M., Delaux, P.M., Li, F.W., Melkonian, B., Mavrodiev, E. V., Sun, W., Fu, Y., Yang, H., Soltis, D.E., Graham, S.W., Soltis, P.S., Liu, X., Xu, X., and Wong, G.K.S., 2018. 10KP: A phylodiverse genome sequencing plan. *GigaScience*. 7 (3), pp. 1–9.

Chong, X., Su, J., Wang, F., Wang, H., Song, A., Guan, Z., Fang, W., Jiang, J., Chen, S., Chen, F., and Zhang, F., 2019. Identification of favorable SNP alleles and candidate genes

responsible for inflorescence-related traits via GWAS in chrysanthemum. *Plant Molecular Biology*. 99 (4–5), pp. 407–420.

Christodoulou, M.D., Battey, N.H., and Culham, A., 2018. Can you make morphometrics work when you know the right answer? Pick and mix approaches for apple identification. *PLOS ONE*. 13 (10): e0205357.

Christodoulou, M.D. and Culham, A., 2018. Looks do matter: the case for morphology in plant description. unpublished.

Clement, M., Posada, D., and Crandall, K.A., 2000. TCS: a computer program to estimate gene genealogies. *Molecular Ecology*. 9, pp. 1657–1659.

Coates, A.M., 1956. Flowers and their histories. London: Hulton Press.

Coissac, E., Boyer, F., Douzet, R., Hombiat, A., Lavergne, S., Perrier, C., Roquet, C., Taberlet, P., and Thiuller, W., 2018. PhyloAlps. Available from: http://phyloalps.org/.

Coissac, E., Hollingsworth, P.M., Lavergne, S., and Taberlet, P., 2016. From barcodes to genomes: Extending the concept of DNA barcodiing. *Molecular Ecology*. 25 (7), pp. 1423–1428.

Corney, D.P.A., Tang, H.L., Clark, J.Y., Hu, Y., and Jin, J., 2012. Automating digital leaf measurement: The tooth, the whole tooth, and nothing but the tooth. *PLOS ONE*. 7 (8), pp. 1–10.

Cronn, R., Knaus, B.J., Liston, A., Maughan, P.J., Parks, M., Syring, J. V., and Udall, J., 2012. Targeted enrichment strategies for next-generation plant biology. *American Journal of Botany*. 99 (2), pp. 291–311.

Cubey, J., 2017. RHS Plant Finder 2017. London: Royal Horticultural Society.

Culham, A. and Grant, M.L., 1999. DNA markers for cultivar identification. In: S. Andrews, A.C. Leslie and C. Alexander, eds. Taxonomy of Cultivated Plants: Third International Symposium. Royal Botanic Gardens, Kew. pp. 183–198.

Cutter, A.D., 2013. Integrating phylogenetics, phylogeography and population genetics through genomes and evolutionary theory. *Molecular Phylogenetics and Evolution*. 69 (3), pp. 1172–1185.

Daffseek, 2018. American Daffodil Society, Inc. [online]. Available from: www.daffseek.org [Accessed 26 Jan 2018].

Davey, J.W., Hohenlohe, P.A., Etter, P.D., Boone, J.Q., Catchen, J.M., and Blaxter, M.L., 2011. Genome-wide genetic marker discovery and genotyping using next-generation sequencing. *Nature Reviews Genetics*. 12 (7), pp. 499–510.

Debiasse, M.B. and Hellberg, M.E., 2015. Discordance between morphological and molecular species boundaries among Caribbean species of the reef sponge *Callyspongia*. *Ecology and Evolution*. 5 (3), pp. 663–675.

De Candolle, A.P., 1816. In: P. J. Redouté ed. Les Liliacées. 8: 126.

Dierckxsens, N., Mardulyn, P., and Smits, G., 2016. NOVOPlasty: De novo assembly of organelle genomes from whole genome data. *Nucleic Acids Research*. 45 (4): e18.

Dodsworth, S., 2015. Genome skimming for next-generation biodiversity analysis. *Trends in Plant Science*. 20 (9), pp. 525–527.

Dong, W., Xu, C., Li, C., Sun, J., Zuo, Y., Shi, S., Cheng, T., Guo, J., and Zhou, S., 2015. *ycf1*, the most promising plastid DNA barcode of land plants. *Scientific Reports*. 5, p. 8348.

Doyen, J.T. and Slobodchikoff, C.N., 1974. An operational approach to species classification. *Systematic Zoology*. 23, pp. 239–247.

Doyle, J.J. and Doyle, J.L., 1987. A rapid DNA isolation procedure for small amounts of fresh leaf tissue. *Phytochemical Bulletin*. 19, pp. 11–15.

Dray, S. and Dufour, A.B., 2007. The ade4 Package: Implementing the Duality Diagram for Ecologists. *Journal of Statistical Software*. 22 (4).

Dubuisson, J.Y., Hébant-Mauri, R., and Galtier, J., 1998. Molecules and Morphology: Conflicts and Congruence within the Fern Genus *Trichomanes* (Hymenophyllaceae). *Molecular Phylogenetics and Evolution*. 9 (3), pp. 390–397.

Dweck, A.C., 2002. The folklore of *Narcissus*. In: G.R. Hanks, ed. *Narcissus* and Daffodil: The genus *Narcissus*. London, UK: Taylor and Francis. pp. 19–29.

Edgar, R.C., 2004. MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research*. 32 (5), pp. 1792–1797.

Feliner, G.N. and Rosselló, J.A., 2007. Better the devil you know? Guidelines for insightful utilization of nrDNA ITS in species-level evolutionary studies in plants. *Molecular Phylogenetics and Evolution*. 44 (2), pp. 911–919.

Fennane, M., Ibn Tattou, M., and El Oualidi, J., 2014. Flore pratique du Maroc. Volume 3. Rabat: Cana Print.

Fernandes, A., 1951. Sur la phylogenie des especes du genre *Narcissus* L. *Boletim Da Sociedade Broteriana*. 25, pp. 113–194.

Fernandes, A., 1968. Keys to the identification of native and naturalised taxa of the genus *Narcissus* L. *R.H.S. Daffodil Year Book*. pp. 37–66.

Fernandes, A., 1975. L'evolution chez le genre *Narcissus* L. *Anales del Instituto Botánico A. J. Cavanilles*. 32 (2), pp. 843–872.

Fernández Casas, J., 1983. Materiales para una monografía de *Narcissus* L. Fontqueria. 3, pp. 23–32.

Fernández Casas, J., 1984a. Remiendos y enmiendas en el género *Narcissus* L. Fontqueria. 6, pp. 35–50.

Fernández Casas, J., 1984b. Dos novedades en *Narcissus* L. Fontqueria. 5, pp. 35–38.

Fernández Casas, J., 1990. Catálogo florístico. *Fontqueria*. 30, p. 30.

Fernández Casas, J., 2005. Narcissorum notulae VI. *Fontqueria*. 55 (34), pp. 265–272.

Fernández Casas, J., 2013. Narcissorum notulae XXXIV. *Fontqueria*. 56 (32), pp. 297–300.

Fernández Casas, J. and Uribe-Echebarría Díaz, P.M., 2011. Narcissorum notulae XXV. *Fontqueria*. 56 (18), pp. 159–164.

Flowers & Plants Association, 2006. Top ten best-selling cult flowers in the UK. Flowers and Plants Association [online]. Available from: http://www.flowersandplantsassociation.org.uk/flowers/trivia/toptenflowers-sales-06.htm [Accessed 8 Nov 2018].

Fonseca, J.P., Levy, A., Henriques, R., Costa, J.C., Neto, C., and Robalo, J., 2015. Phylogenenetic approach of the section *Bulbocodii* D.C. of *Narcissus* based on cpDNA. A case of taxonomic inflation? *Plant Biosystems.* pp. 1–12.

Ford, C.S., Ayres, K.L., Toomey, N., Haider, N., Van Alphen Stahl, J., Kelly, L.J., Wikström, N., Hollingsworth, P.M., Duff, R.J., Hoot, S.B., Cowan, R.S., Chase, M.W., and Wilkinson, M.J., 2009. Selection of candidate coding DNA barcoding regions for use on land plants. *Botanical Journal of the Linnean Society*. 159 (1), pp. 1–11.

Galli, Z., Halasz, G., Kiss, E., Heszky, L., and Dobránszki, J., 2005. Molecular Identification of Commercial Apple Cultivars with Microsatellite Markers. *HortScience*. 40 (7), pp. 1974–1977.

Garnett, S.T. and Christidis, L., 2017. Taxonomy anarchy hampers conservation. *Nature*. 546, pp. 25–27.

GBIF.org, 2018. GBIF Occurence Download [online]. Available from: https://www.gbif.org/occurrence/search?q=*Narcissus* [Accessed 1 Dec 2018].

Gilmore, B., Bassil, N., Nyberg, A., Knaus, B., Smith, D., Barney, D.L., and Hummer, K., 2013. Microsatellite Marker Development in Peony using Next Generation Sequencing. *Journal of the American Society for Horticultural Science*. 138 (1), pp. 64–74.

Givnish, T.J., Zuluaga, A., Spalink, D., Soto Gomez, M., Lam, V.K.Y., Saarela, J.M., Sass, C., Iles, W.J.D., De Sousa, D.J.L., Leebens-Mack, J., Chris Pires, J., Zomlefer, W.B., Gandolfo, M.A., Davis, J.I., Stevenson, D.W., DePamphilis, C., Specht, C.D., Graham, S.W., Barrett, C.F., and Ané, C., 2018. Monocot plastid phylogenomics, timeline, net rates of species diversification, the power of multi-gene analyses, and a functional model for the origin of monocots. *American Journal of Botany*. 105 (11), pp. 1888–1910.

Goodwin, S., McPherson, J.D., and McCombie, W.R., 2016. Coming of age: ten years of next-generation sequencing technologies. *Nature Reviews Genetics.* 17 (6), pp. 333–351.

Graham, S.W. and Barrett, S.C.H., 2004. Phylogenetic reconstruction of the evolution of stylar polymorphism in *Narcissus* (Amaryllidaceae). *American Journal of Botany*. 91 (7), pp. 1007–1021.

Guo, Y.Y., Huang, L.Q., Liu, Z.J., and Wang, X.Q., 2016. Promise and challenge of DNA barcoding in Venus slipper (*Paphiopedilum*). *PLOS ONE*. 11 (1), p. e0146880.

Hanks, G.R., 2002a. Commercial production of *Narcissus* bulbs. In: *Narcissus* and Daffodil: The genus *Narcissus*. London, UK: Taylor and Francis. pp. 53–130.

Hanks, G.R., 2002b. *Narcissus* and Daffodil. London: Taylor and Francis.

Harrison, N., Harrison, R.J., and Kidner, C. a., 2016. Comparative analysis of *Begonia* plastid genomes and their utility for species-level phylogenetics. *PLOS ONE*. 11 (4), p. e0153248.

Haworth, A.H., 1831. Narcissinearum monographia. Edition 2. London: Ridgway.

Heap, M.J., Culham, A., Lenoir, J., and Gavilán, R.G., 2014. Can the Iberian floristic diversity withstand near-future climate change? *Open Journal of Ecology*. 4 (17), pp. 1089–1101.

Heuertz, M., Carnevale, S., Fineschi, S., Sebastiani, F., Hausman, J.F., Paule, L., and Vendramin, G.G., 2006. Chloroplast DNA phylogeography of European ashes, *Fraxinus* sp. (Oleaceae): Roles of hybridization and life history traits. *Molecular Ecology*. 15 (8), pp. 2131–2140.

Hey, J., Waples, R.S., Arnold, M.L., Butlin, R.K., and Harrison, R.G., 2003. Understanding and confronting species uncertainty in biology and conservation. *Trends in Ecology and Evolution*. 18 (11), pp. 597–603.

Heywood, V.H., Brummit, R.K., Culham, A., and Seberg, O., 2007. Amaryllidaceae. In: Flowering Plant Families of the World. Ontario, Canada: Firefly Books.

Hodgins, K.A. and Barrett, S.C.H., 2007. Population structure and genetic diversity in tristylous *Narcissus triandrus*: Insights from microsatellite and chloroplast DNA variation. *Molecular Ecology*. 16 (11), pp. 2317–2332.

Hodgins, K.A., Stehlik, I., Wang, P., and Barrett, S.C.H., 2007. The development of eight microsatellite loci in the wild daffodil *Narcissus triandrus* (Amaryllidaceae): Primer note. *Molecular Ecology Notes*. 7 (3), pp. 510–512.

Hollingsworth, P.M., Graham, S.W., and Little, D.P., 2011. Choosing and using a plant DNA barcode. *PLOS ONE.* 6 (5), p. e19254.

Hughes, C. and Eastwood, R., 2006. Island radiation on a continental scale: exceptional rates of plant diversification after uplift of the Andes. *Proceedings of the National Academy of Sciences of the United States of America*. 103 (27), pp. 10334–10339.

Jang, C.G., Müllner, A.N., and Greimler, J., 2005. Conflicting patterns of genetic and morphological variation in European *Gentianella* section *Gentianella*. *Botanical Journal of the Linnean Society*. 148 (2), pp. 175–187.

Jha, S. and Dick, C.W., 2010. Native bees mediate long-distance pollen dispersal in a shade coffee landscape mosaic. *Proceedings of the National Academy of Sciences*. 107 (31), pp. 13760–13764.

Jiménez, J.F., Sánchez-Gómez, P., Guerra, J., Molins, A., and Rosselló, J.A., 2009. Regional speciation or taxonomic inflation? The status of several narrowly distributed and

endangered species of narcissus using ISSR and nuclear ribosomal ITS markers. *Folia Geobotanica*. 44 (2), pp. 145–158.

Jombart, T., 2008. adegenet: a *R* package for the multivariate analysis of genetic markers. *Bioinformatics*. 24 (11), pp. 1403–1405.

Jombart, T. and Collins, C., 2015. A tutorial for Discriminant Analysis of Principal Components (DAPC) using adegenet 2.0.0 [online]. Available from: http://adegenet.r-forge.r-project.org/files/tutorial-dapc.pdf.

Katoh, K. and Standley, D.M., 2013. MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Molecular Biology and Evolution*. 30 (4), pp. 772–780.

Kearse, M., Moir, R., Wilson, A., Stone-Havas, S., Cheung, M., Sturrock, S., Buxton, S., Cooper, A., Markowitz, S., Duran, C., Thierer, T., Ashton, B., Mentjies, P., and Drummond, A., 2012. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics*. 28 (12), pp. 1647–1649.

Kim, Y.M., Kim, S., Koo, N., Shin, A.Y., Yeom, S.I., Seo, E., Park, S.J., Kang, W.H., Kim, M.S., Park, J., Jang, I., Kim, P.G., Byeon, I., Kim, M.S., Choi, J.H., Ko, G., Hwang, J.H., Yang, T.J., Choi, S.B., Lee, J.M., Lim, K.B., Lee, J., Choi, I.Y., Park, B.S., Kwon, S.Y., Choi, D., and Kim, R.W., 2017. Genome analysis of *Hibiscus syriacus* provides insights of polyploidization and indeterminate flowering in woody plants. *DNA research*. 24 (1), pp. 71–80.

Kington, S., 2008. Daffodil Register and Classified List 2008. London: Royal Horticultural Society.

Kjos, M., Fjellheim, S., Rognli, O.A., and Hvoslef-Eide, A.K., 2010. Amplified fragment length polymorphism (AFLP) markers for fingerprinting of *Argyranthemum frutescens* cultivars. *Scientia Horticulturae*. 124 (4), pp. 506–510.

Könyves, K., 2014. Taxonomy to underpin cultivar identification in hoop-petticoat daffodils. PhD thesis. University of Reading.

Könyves, K., Bilsborrow, J., David, J., and Culham, A., 2018. The complete chloroplast genome of *Narcissus poeticus* L. (Amaryllidaceae: Amaryllidoideae). *Mitochondrial DNA Part B*. 3 (2), pp. 1137–1138.

Könyves, K. and Culham, A., 2014. *C. colchicum* - variation measured by DNA sequence. *Cyclamen*. 38 (1), pp. 18–21.

Könyves, K., David, J., and Culham, A., 2019. Jumping through the hoops: the challenges of daffodil (*Narcissus*) classification. *Botanical Journal of the Linnean Society*. 190, pp. 389–404.

Könyves, K., David, J.C., and Culham, A., 2011. Towards a reproducible system of identification for daffodil cultivars. *Daffodil, Snowdrop and Tulip Year Book*. pp. 26–29.

Könyves, K., David, J.C., and Culham, A., 2016. Microsatellite Markers for Hoop-Petticoat Daffodils (*Narcissus* sect. *Bulbocodii*, Amaryllidaceae). *Applications in Plant Sciences*. 4 (4): 1500127.

Kostamo, K., Toljamo, A., Antonius, K., Kokko, H., and Kärenlampi, S.O., 2013. Morphological and molecular identification to secure cultivar maintenance and management of self-sterile *Rubus arcticus*. *Annals of Botany*. 111 (4), pp. 713–721.

Kramina, T.E., Meschersky, I.G., Degtjareva, G. V., Samigullin, T.H., Belokon, Y.S., and Schanzer, I.A., 2018. Genetic variation in the *Lotus corniculatus* complex (Fabaceae) in northern Eurasia as inferred from nuclear microsatellites and plastid *trnL-trnF* sequences. *Botanical Journal of the Linnean Society*. 188, pp. 87–116.

Kyriakidou, M., Tai, H.H., Anglin, N.L., Ellis, D., and Strömvik, M. V., 2018. Current Strategies of Polyploid Plant Genome Sequence Assembly. *Frontiers in Plant Science*. 9, pp. 1–15.

Larsson, A., 2014. Aliview: a fast and lightweight alignment viewer and editor for large data sets. *Bioinformatics*. 30 (22), pp. 3276–3278.

Lawson, D.J., Van Dorp, L., and Falush, D., 2018. A tutorial on how not to over-interpret STRUCTURE and ADMIXTURE bar plots. *Nature Communications*. 9 (1), pp. 1–11.

Li, C., Lin, F., An, D., Wang, W., and Huang, R., 2017. Genome Sequencing and Assembly by Long Reads in Plants. *Genes*. 9 (1), p. 6.

Li, F.-W. and Harkess, A., 2018. A guide to sequence your favorite plant genomes. *Applications in Plant Sciences*. 6 (3), p. e1030.

Li, X., Yang, Y., Henry, R.J., Rossetto, M., Wang, Y., and Chen, S., 2015. Plant DNA barcoding: from gene to genome. *Biological reviews*. 90 (1), pp. 157–166.

Li, Y., Feng, Y., Wang, X.Y., Liu, B., and Lv, G.H., 2014. Failure of DNA barcoding in discriminating *Calligonum* species. *Nordic Journal of Botany*. 32 (4), pp. 511–517.

Li, Y., Zhang, Z., Yang, J., and Lv, G., 2018. Complete chloroplast genome of seven *Fritillaria* species, variable DNA markers identification and phylogenetic relationships within the genus. *PLOS ONE*. 13 (3).

LLoret i Sabaté, F.J. and Fernández Casas, J., 2009. *Narcissus* × *cardonae* (Amaryllidaceae). Nuevo mesto silvestre. *Fontqueria*. 56 (1), pp. 1–16.

Lobel, M., 1570. Stirpium Adversaria Nova. London.

Lomba, A., Pellissier, L., Randin, C., Vicente, J., Moreira, F., Honrado, J., and Guisan, A., 2010. Overcoming the rare species modelling paradox: A novel hierarchical framework applied to an Iberian endemic plant. *Biological Conservation*. 143 (11), pp. 2647–2657.

Lorda López, M., 2013. Catálogo florístico de Navarra. Pamplona, Spain: Jolube Cinsultor y Editor Botánico.

Lu, J., Liu, Y., Xu, J., Mei, Z., Shi, Y., Liu, P., He, J., Wang, X., Meng, Y., Feng, S., Shen, C., and Wang, H., 2018. High-density genetic map construction and stem total polysaccharide content-related QTL exploration for Chinese endemic *Dendrobium* (Orchidaceae). *Frontiers in Plant Science*. 9, pp. 1–11.

Lu, R.-S., Li, P., and Qiu, Y.-X., 2017. The complete chloroplast genomes of three *Cardiocrinum* (Liliaceae) species: comparative genomic and phylogenetic analyses. *Frontiers in Plant Science*. 7, pp. 1–12.

Makki, R.M., Saeedi, A.A., Khan, T.K., Ali, H.M., and Ramadan, A.M., 2019. Single nucleotide polymorphism analysis in plastomes of eight *Catharanthus roseus* cultivars. *Biotechnology & Biotechnological Equipment*. 33 (1), pp. 419–428.

MarketInsider, 2015. Daffodil industry in UK marks the arrival of spring. International Trade Centre [online]. Available from: http://www.intracen.org/blog/Daffodil-industry-in-UK-marks-the-arrival-of-spring/ [Accessed 15 Apr 2018].

Marques, I., Aguilar, J.F., Martins-Louçao, M.A., Moharrek, F., and Feliner, G.N., 2017. A three-genome five-gene comprehensive phylogeny of the bulbous genus *Narcissus* (Amaryllidaceae) challenges current classifications and reveals multiple hybridization events. *Taxon*. 66 (4), pp. 832–854.

Marques, I., Feliner, G.N., Munt, D.D., Martins-Loução, M.A., and Aguilar, J.F., 2010. Unraveling cryptic reticulate relationships and the origin of orphan hybrid disjunct populations in *Narcissus*. *Evolution*. 64 (8), pp. 2353–2368.

Marques, I., Jürgens, A., Aguilar, J.F., and Feliner, G.N., 2016. Convergent recruitment of new pollinators is triggered by independent hybridization events in *Narcissus*. *New Phytologist*. 210 (2), pp. 731–742.

Marques, I., Loureiro, J., Draper, D., Castro, M., and Castro, S., 2018. How much do we know about the frequency of hybridisation and polyploidy in the Mediterranean region? *Plant Biology*. 20, pp. 21–37.

Marques, I., Nieto Feliner, G., Martins-Loução, M.A., and Fuertes Aguilar, J., 2012. Genome size and base composition variation in natural and experimental *Narcissus* (Amaryllidaceae) hybrids. *Annals of Botany.* 109 (1), pp. 257–264.

Marschall, T., Marz, M., Abeel, T., Dijkstra, L., Dutilh, B.E., Ghaffaari, A., Kersey, P., Kloosterman, W.P., Mäkinen, V., Novak, A.M., Paten, B., Porubsky, D., Rivals, E., Alkan, C., Baaijens, J.A., De Bakker, P.I.W., Boeva, V., Bonnal, R.J.P., Chiaromonte, F., Chikhi, R., Ciccarelli, F.D., Cijvat, R., Datema, E., Van Duijn, C.M., Eichler, E.E., Ernst, C., Eskin, E., Garrison, E., El-Kebir, M., Klau, G.W., Korbel, J.O., Lameijer, E.W., Langmead, B., Martin, M., Medvedev, P., Mu, J.C., Neerincx, P., Ouwens, K., Peterlongo, P., Pisanti, N., Rahmann, S., Raphael, B., Reinert, K., De Ridder, D., De Ridder, J., Schlesner, M., Schulz-Trieglaff, O., Sanders, A.D., Sheikhizadeh, S., Shneider, C., Smit, S., Valenzuela, D., Wang, J., Wessels, L., Zhang, Y., Guryev, V., Vandin, F., Ye, K., and Schönhuth, A., 2016. Computational pan-genomics: status, promises and challenges. *Briefings in Bioinformatics*. 19 (1), p. bbw089.

Matasci, N., Hung, L.H., Yan, Z., Carpenter, E.J., Wickett, N.J., Mirarab, S., Nguyen, N., Warnow, T., Ayyampalayam, S., Barker, M., Burleigh, J.G., Gitzendanner, M.A., Wafula, E., Der, J.P., DePamphilis, C.W., Roure, B., Philippe, H., Ruhfel, B.R., Miles, N.W., Graham, S.W., Mathews, S., Surek, B., Melkonian, M., Soltis, D.E., Soltis, P.S., Rothfels, C., Pokorny, L., Shaw, J.A., DeGironimo, L., Stevenson, D.W., Villarreal, J.C., Chen, T., Kutchan, T.M., Rolf, M., Baucom, R.S., Deyholos, M.K., Samudrala, R., Tian, Z., Wu, X., Sun, X., Zhang, Y., Wang, J., Leebens-Mack, J., and Wong, G.K.S., 2014. Data access for the 1,000 Plants (1KP) project. *GigaScience*. 3 (1), pp. 1–10.

Mathew, B., 2002. Classification of the genus *Narcissus*. In: *Narcissus* and Daffodil: The genus *Narcissus*. pp. 30–52.

Mayr, G., 2011. The phylogeny of charadriiform birds (shorebirds and allies) - reassessing the conflict between morphology and molecules. *Zoological Journal of the Linnean Society*. 161 (4), pp. 916–934.

McClaren, B., 2009. Encyclopedia of dahlias. In: M. McClaren, ed. *Dahlia*: history and species. Portland: Timber Press. pp. 161–166.

McKain M.R., and Wilson M., 2017. Fast-Plast: Rapid de novo assembly and finishing for whole chloroplast genomes. Version 1.2.6.

McKain M.R., Hartsock R.H., Wohl M.M., and Kellogg E.A., 2017. Verdant: Automated annotation, alignment and phylogenetic analysis of whole chloroplast genomes. *Bioinformatics*. 33:130-132.

Medrano, M., López-Perea, E., and Herrera, C.M., 2014. Population Genetics Methods Applied to a Species Delimitation Problem: Endemic Trumpet Daffodils (*Narcissus* Section *Pseudonarcissi*) from the Southern Iberian Peninsula. *International Journal of Plant Sciences*. 175 (5), pp. 501–517.

Merino, B., 1909. Flora descriptiva é illustrada de Galicia. Santiago.

Miah, G., Rafii, M.Y., Ismail, M.R., Puteh, A.B., Rahim, H.A., Islam, N.K., and Latif, M.A., 2013. A review of microsatellite markers and their applications in rice breeding programs to improve blast disease resistance. *International Journal of Molecular Sciences*. 14 (11), pp. 22499–22528.

Miryeganeh, M., Takayama, K., Tateishi, Y., and Kajita, T., 2014. Long-distance dispersal by sea-drifted seeds has maintained the global distribution of *Ipomoea pes-caprae* subsp. *brasiliensis* (Convolvulaceae). *PLOS ONE*. 9 (4): e91836.

Moritz, C. and Cicero, C., 2004. DNA barcoding: Promise and pitfalls. *PLOS Biology*. 2 (10) e354.

Moriya, S., Iwanami, H., Okada, K., Yamamoto, T., and Abe, K., 2011. A practical method for apple cultivar identification and parent-offspring analysis using simple sequence repeat markers. *Euphytica*. 177 (1), pp. 135–150.

Muñoz-Rodríguez, P., Carruthers, T., Wood, J.R.I., Williams, B.R.M., Weitemier, K., Kronmiller, B., Ellis, D., Anglin, N.L., Longway, L., Harris, S.A., Rausher, M.D., Kelly, S., Liston, A., and Scotland, R.W., 2018. Reconciling Conflicting Phylogenies in the Origin of Sweet Potato and Dispersal to Polynesia. *Current Biology*. 28 (8), pp. 1246-1256.e12.

Nagamitsu, T., 2017. Genetic structure in chloroplast and nuclear microsatellites in *Rosa rugosa* around sea straits in northern Japan. *Plant Species Biology*. 32, pp. 359–367.

Navarro, L., Ayensa, G., Ferrero, V., and Sánchez, J.M., 2012. The avoidance of self-interference in the endemic daffodil *Narcissus cyclamineus* (Amaryllidaceae). *Plant Ecology*. 213 (11), pp. 1813–1822.

Nowoshilow, S., Schloissnig, S., Fei, J.-F., Dahl, A., Pang, A.W.C., Pippel, M., Winkler, S., Hastie, A.R., Young, G., Roscito, J.G., Falcon, F., Knapp, D., Powell, S., Cruz, A., Cao, H., Habermann, B., Hiller, M., Tanaka, E.M., and Myers, E.W., 2018. The axolotl genome and the evolution of key tissue formation regulators. *Nature*. 554 (7690), pp. 50–55.

Nylander, J.A.A., 2004. MrModeltest2. Available from: https://github.com/nylander/MrModeltest2.

De Oliveira Plotze, R. and Bruno, O.M., 2009. Automatic leaf structure biometry: Computer vision techniques and their applications in plant taxonomy. *International Journal of Pattern Recognition and Artificial Intelligence*. 23 (02), pp. 247–262.

Oxford Economics, 2018. The Economic Impact of Ornamental Horticulture and Landscaping in the UK.

Oxford Nanopore Technologies, 2018. Closing the gap in plant genomes.

Van de Paer, C., Bouchez, O., and Besnard, G., 2017. Prospects on the evolutionary mitogenomics of plants: A case study on the olive family (Oleaceae). *Molecular Ecology Resources*. 18 (3), pp. 407–423.

Pagel, M., Meade, A., and Barker, D., 2004. Bayesian estimation of ancestral character states on phylogenies. *Systematic Biology*. 53 (5), pp. 673–684.

Paradis, E. and Schliep, K., 2019. ape 5.0: an environment for modern phylogenetics and evolutionary analyses in *R*. *Bioinformatics*. 35 (3), pp. 526–528.

Van de Passe, C., 1614. Hortus Floridus. Extant Arnhemii.

Peakall, R. and Smouse, P.E., 2012. GenALEx 6.5: Genetic analysis in Excel. Population genetic software for teaching and research-an update. *Bioinformatics*. 28 (19), pp. 2537–2539.

Peltier, D., Becher, S.A., Steinmetz, K., Weising, K., Boury, S., Renou, J.P., Kahl, G., and Wolff, K., 2000. Microsatellites for cultivar identification in *Pelargonium*. *Theoretical and Applied Genetics*. (101), pp. 643–651.

Percy, D.M., Argus, G.W., Cronk, Q.C., Fazekas, A.J., Kesanakurti, P.R., Burgess, K.S., Husband, B.C., Newmaster, S.G., Barrett, S.C.H., and Graham, S.W., 2014. Understanding the spectacular failure of DNA barcoding in willows (*Salix*): Does this result from a trans-specific selective sweep? *Molecular Ecology*. 23 (19), pp. 4737–4756.

Pérez-Barrales, R., Arroyo, J., and Armbruster, W.S., 2007. Differences in pollinator faunas may generate geographic differences in floral morphology and integration in *Narcissus papyraceus* (Amaryllidaceae). *Oikos*. 116, pp. 1904–1918.

Pérez-Barrales, R., Pino, R., Albaladejo, R.G., and Arroyo, J., 2009. Geographic variation of flower traits in *Narcissus papyraceus* (Amaryllidaceae): Do pollinators matter? *Journal of Biogeography*. 36 (7), pp. 1411–1422.

Pérez-Barrales, R., Vargas, P., and Arroyo, J., 2004. Convergent evolution of flower polymorphism in *Narcissus* (Amaryllidaceae). *New Phytologist*. 161 (1), pp. 235–252.

Pérez-Barrales, R., Vargas, P., and Arroyo, J., 2006. New evidence for the Darwinian hypothesis of heterostyly: Breeding systems and pollinators in *Narcissus* sect. *Apodanthi*. *New Phytologist*. 171 (3), pp. 553–567.

Pérez, R., Vargas, P., and Arroyo, J., 2003. Convergent evolution of flower polymorphism in *Narcissus* (Amaryllidaceae). *New Phytologist*. 161 (1), pp. 235–252.

Pfeiffer, T., Roschanski, A.M., Pannell, J.R., Korbecka, G., and Schnittler, M., 2011. Characterization of microsatellite loci and reliable genotyping in a polyploid plant, *Mercurialis perennis* (Euphorbiaceae). *Journal of Heredity*. 102 (4), pp. 479–488.

Pugsley, H.W., 1933. Monograph of *Narcissus*, subgenus *Ajax*. Journal of the Royal Horticultural Society. 58, pp. 17–93.

Raman, G. and Park, S., 2016. The complete chloroplast genome sequence of *Ampelopsis*: gene organization, comparative analysis, and phylogenetic relationships to other Angiosperms. *Frontiers in Plant Science*. 7, pp. 1–13.

Rambaut, A., Drummond, A.J., Xie, D., Baele, G., and Suchard, M.A., 2018. Posterior summarization in Bayesian phylogenetics using Tracer 1.7. *Systematic biology*. 67 (5), pp. 901–904.

Ripma, L.A., Simpson, M.G., and Hasenstab-Lehman, K., 2014. Geneious! Simplified genome skimming methods for phylogenetic systematic studies: A case study in *Oreocarya* (Boraginaceae). *Applications in Plant Sciences*. 2 (12), p. 1400062.

Rivera Nuñez, D., Obón De Castro, C., Ríos Ruiz, S., and Alcaraz Ariza, F., 2003. The origin of cultivation and wild ancestors of daffodils (*Narcissus* subgenus *Ajax*) (Amaryllidaceae) from an analysis of early illustrations. *Scientia Horticulturae*. 98 (4), pp. 307–330.

Ronquist, F., Teslenko, M., Van Der Mark, P., Ayres, D.L., Darling, A., Höhna, S., Larget, B., Liu, L., Suchard, M.A., and Huelsenbeck, J.P., 2012. Mrbayes 3.2: Efficient bayesian phylogenetic inference and model choice across a large model space. *Systematic Biology*. 61 (3), pp. 539–542.

Ross, T.G., Barrett, C.F., Soto Gomez, M., Lam, V.K.Y.Y., Henriquez, C.L., Les, D.H., Davis, J.I., Cuenca, A., Petersen, G., Seberg, O., Thadeo, M., Givnish, T.J., Conran, J., Stevenson, D.W., and Graham, S.W., 2016. Plastid phylogenomics and molecular evolution of Alismatales. *Cladistics*. 32 (2), pp. 160–178.

Salih, R.H.M., Majeský, L., Schwarzacher, T., Gornall, R., and Heslop-Harrison, P., 2017. Complete chloroplast genomes from apomictic *Taraxacum* (Asteraceae): Identity and variation between three microspecies. *PLOS ONE*. 12 (2), pp. 1–22.

Santos-Gally, R., Gonzalez-Voyer, A., and Arroyo, J., 2013. Deconstructing heterostyly: The evolutionary role of incompatibility system, pollinators, and floral architecture. *Evolution*. 67 (7), pp. 2072–2082.

Santos-Gally, R., Vargas, P., and Arroyo, J., 2012. Insights into Neogene Mediterranean biogeography based on phylogenetic relationships of mountain and lowland lineages of *Narcissus* (Amaryllidaceae). *Journal of Biogeography*. 39 (4), pp. 782–798.

Shaffer, H.B. and Thomson, R.C., 2007. Delimiting species in recent radiations. *Systematic Biology*. 56 (6), pp. 896–906.

Seberg O., Petersen G., Davis J.I., Pires J.C., Stevenson D.W., Chase M.W., Fay M.F., Devey D.S., Jørgensen T., Sytsma K.J., and Pillon Y., 2012. Phylogeny of the Asparagales based on three plastid and two mitochondrial genes. *American Journal of Botany*. 99: pp. 875-889.

Simon-Porcar, V.I., Pico, F.X., and Arroyo, J., 2015. Range-wide population genetics and variation in morph ratio in style-dimorphic *Narcissus papyraceus*. *American Journal of Botany*. 102 (3), pp. 449–456.

Simón, V.I., Xavier Picó, F., and Arroyo, J., 2010. New microsatellite loci for *Narcissus papyraceus* (Amarillydaceae) and cross-amplification in other congeneric species. *American Journal of Botany.* 97 (3), pp. 10–13.

Sinama, M., Dubut, V., Costedoat, C., Gilles, A., Junker, M., Malausa, T., Martin, J.F., Nève, G., Pech, N., Schmitt, T., Zimmermann, M., and Meglécz, E., 2011a. Challenges of microsatellite development in Lepidoptera: *Euphydryas aurinia* (Nymphalidae) as a case study. *European Journal of Entomology*. 108 (2), pp. 261–266.

Sitther, V., Road, W., Harris, D.L., Valley, F., and Okie, W.R., 2012. Cultivar Identification , Pedigree Verification , and Diversity Analysis among Peach Cultivars Based on Simple Sequence Repeat Markers. *Journal of American Society of Horticultural Science*. 137 (2), pp. 114–121.

Sotiaux, A., Enroth, J., Olsson, S., Quandt, D., and Vanderpoorten, A., 2009. When morphology and molecules tell us different stories: a case-in-point with *Leptodon corsicus*, a new and unique endemic moss species from Corsica. *Journal of Bryology*. 31 (3), pp. 186–196.

Souza Silva, M.F., Moreira de Andrade, I., and Joseph Mayo, S., 2012. Geometric morphometrics of leaf blade shape in *Montrichardia linifera* (Araceae) populations from the Rio Parnaíba Delta, north-east Brazil. *Botanical Journal of the Linnean Society*. 170, pp. 554–572.

Spach, E., 1846. Histoire Naturelle des Végétaux: Phanérogames. volume 12. Paris: Librairie Encyclopedique de Roret.

Stallman, J.K., Funk, V.A., Price, J.P., and Knope, M.L., 2019. DNA barcodes fail to accurately differentiate species in Hawaiian plant lineages. *Botanical Journal of the Linnean Society*. 190 (4), pp. 374–388.

Stamatakis A., 2014. RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*. 30: pp. 1312-1313.

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

Terry, R.G., Brown, G.K., and Olmstead, R.G., 1997. Examination of subfamilial phylogeny in Bromeliaceae using comparative sequencing of the plastid locus *ndhF*. *American Journal of Botany*. 84 (5), pp. 664–670.

Thompson, J.D., 2005. Variation and evolution of reproductive traits in the Mediterranean mosaic. In: Plant Evolution in the Mediterranean. pp. 167–207.

Thomson, S.A., Pyle, R.L., Ahyong, S.T., Alonso-Zarazaga, M., Ammirati, J., Araya, J.F., Ascher, J.S., Audisio, T.L., Azevedo-Santos, V.M., Bailly, N., Baker, W.J., Balke, M., Barclay, M.V.L., Barrett, R.L., Benine, R.C., Bickerstaff, J.R.M., Bouchard, P., Bour, R., Bourgoin, T., Boyko, C.B., Breure, A.S.H., Brothers, D.J., Byng, J.W., Campbell, D., Ceríaco, L.M.P., Cernák, I., Cerretti, P., Chang, C.-H., Cho, S., Copus, J.M., Costello, M.J., Cseh, A., Csuzdi, C., Culham, A., D'Elía, G., D'Udekem d'Acoz, C., Daneliya, M.E., Dekker, R., Dickinson, E.C.,

Dickinson, T.A., Van Dijk, P.P., Dijkstra, K.-D.B., Dima, B., Dmitriev, D.A., Duistermaat, L., Dumbacher, J.P., Eiserhardt, W.L., Ekrem, T., Evenhuis, N.L., Faille, A., Fernández-Triana, J.L., Fiesler, E., Fishbein, M., Fordham, B.G., Freitas, A.V.L., Friol, N.R., Fritz, U., Frøslev, T., Funk, V.A., Gaimari, S.D., Garbino, G.S.T., Garraffoni, A.R.S., Geml, J., Gill, A.C., Gray, A., Grazziotin, F.G., Greenslade, P., Gutiérrez, E.E., Harvey, M.S., Hazevoet, C.J., He, K., He, X., Helfer, S., Helgen, K.M., Van Heteren, A.H., Hita Garcia, F., Holstein, N., Horváth, M.K., Hovenkamp, P.H., Hwang, W.S., Hyvönen, J., Islam, M.B., Iverson, J.B., Ivie, M.A., Jaafar, Z., Jackson, M.D., Jayat, J.P., Johnson, N.F., Kaiser, H., Klitgård, B.B., Knapp, D.G., Kojima, J.-I., Kõljalg, U., Kontschán, J., Krell, F.-T., Krisai-Greilhuber, I., Kullander, S., Latella, L., Lattke, J.E., Lencioni, V., Lewis, G.P., Lhano, M.G., Lujan, N.K., Luksenburg, J.A., Mariaux, J., Marinho-Filho, J., Marshall, C.J., Mate, J.F., McDonough, M.M., Michel, E., Miranda, V.F.O., Mitroiu, M.-D., Molinari, J., Monks, S., Moore, A.J., Moratelli, R., Murányi, D., Nakano, T., Nikolaeva, S., Noyes, J., Ohl, M., Oleas, N.H., Orrell, T., Páll-Gergely, B., Pape, T., Papp, V., Parenti, L.R., Patterson, D., Pavlinov, I.Y., Pine, R.H., Poczai, P., Prado, J., Prathapan, D., Rabeler, R.K., Randall, J.E., Rheindt, F.E., Rhodin, A.G.J., Rodríguez, S.M., Rogers, D.C., Roque, F.D.O., Rowe, K.C., Ruedas, L.A., Salazar-Bravo, J., Salvador, R.B., Sangster, G., Sarmiento, C.E., Schigel, D.S., Schmidt, S., Schueler, F.W., Segers, H., Snow, N., Souza-Dias, P.G.B., Stals, R., Stenroos, S., Stone, R.D., Sturm, C.F., Štys, P., Teta, P., Thomas, D.C., Timm, R.M., Tindall, B.J., Todd, J.A., Triebel, D., Valdecasas, A.G., Vizzini, A., Vorontsova, M.S., De Vos, J.M., Wagner, P., Watling, L., Weakley, A., Welter-Schultes, F., Whitmore, D., Wilding, N., Will, K., Williams, J., Wilson, K., Winston, J.E., Wüster, W., Yanega, D., Yeates, D.K., Zaher, H., Zhang, G., Zhang, Z.-Q., and Zhou, H.-Z., 2018. Taxonomy based on science is necessary for global conservation. *PLOS* Biology. 16 (3), p. e2005075.

Turland, N.J., Wiersema, J.H., Barrie, F.R., Greuter, W., Hawksworth, D.L., Herendeen, P.S., Knapp, S., Kusber, W.H., Li, D.Z., Markhold, K., May, T.W., McNeill, J., Monro, A.M., Prado, J., Price, M.J., and Smith, G.F., 2018. International Code of Nomenclature for algae, fungi, and plants (Shenzhen Code) adopted by the Nineteenth International Botanical Congress Shenzhen, China, July 2017. In: Regnum Vegetabile. Glashütten: Koeltz Botanical Books.

Twyford, A.D. and Ness, R.W., 2017. Strategies for complete plastid genome sequencing. *Molecular Ecology Resources*. 17 (5), pp. 858–868.

Unamba, C.I.N., Nag, A., and Sharma, R.K., 2015. Next Generation Sequencing technologies: The doorway to the unexplored genomics of non-model plants. *Frontiers in Plant Science*. 6, 1074.

Uribe-Convers, S., Settles, M.L., and Tank, D.C., 2016. A phylogenomic approach based on PCR target enrichment and high throughput sequencing: Resolving the diversity within the south American species of *Bartsia* L. (Orobanchaceae). *PLOS ONE*. 11 (2), p. e0148203.

De Valk, H.A., Meis, J.F.G.M., Bretagne, S., Costa, J.-M., Lasker, B.A., Balajee, S.A., Pasqualotto, A.C., Anderson, M.J., Alcázar-Fuoli, L., Mellado, E., and Klaassen, C.H.W., 2009. Interlaboratory reproducibility of a microsatellite-based typing assay for *Aspergillus fumigatus* through the use of allelic ladders: proof of concept. *Clinical Microbiology and Infection*. 15 (2), pp. 180–187.

Vallet, P., 1608. Le Jardin du Roy très Chrestien Henry IV Roy de France et de Navare. Paris.

Valtueña, F.J., Preston, C.D., and Kadereit, J.W., 2012. Phylogeography of a Tertiary relict plant, *Meconopsis cambrica* (Papaveraceae), implies the existence of northern refugia for a temperate herb. *Molecular Ecology*. 21 (6), pp. 1423–1437.

Vargas, O.M., Ortiz, E.M., and Simpson, B.B., 2017. Conflicting phylogenomic signals reveal a pattern of reticulate evolution in a recent high-Andean diversification (Asteraceae: Astereae: *Diplostephium*). *New Phytologist*. 214 (4), pp. 1736–1750.

Vavrek, M.J., 2011. fossil: palaeoecological and palaeogeographical analysis tools. *Palaeontologia Electronica*. 14:1.

Wang, J.Y. and Chuang, K.C., 2013. Development of novel microsatellite markers for effective applications in *Anthurium* cultivar identification. *Euphytica*. 189 (3), pp. 421–431.

Wang, W., Schalamun, M., Morales-Suarez, A., Kainer, D., Schwessinger, B., and Lanfear, R., 2018. Assembly of chloroplast genomes with long- and short-read data: a comparison of approaches using *Eucalyptus pauciflora* as a test case. *BMC Genomics*. 19 (1), p. 977.

Webb, D.A., 1978. Taxonomic notes on *Narcissus* L. *Botanical Journal of the Linnean Society*. 76 (4), pp. 298–307.

Webb, D.A., 1980. *Narcissus* L. In: T.G. Tutin, V.H. Heywood, N.A. Burges, D.M. Moore, D.H. Valentine, S.M. Walters and D.A. Webb, eds. Flora Europaea. Cambridge: Cambridge University Press. pp. 78–84.

White, O.W., Doo, B., Carine, M.A., and Chapman, M.A., 2016a. Transcriptome Sequencing and Simple Sequence Repeat Marker Development for Three Macaronesian Endemic Plant Species. *Applications in Plant Sciences*. 4 (8), p. 1600050.

Wylie, A.P., 1952. The history of the garden Narcissi. Heredity. 6, pp. 137–156.

Xie, C., Xie, D. feng, Zhong, Y., Guo, X.L., Liu, Q., Zhou, S.D., and He, X.J., 2019. The effect of Hengduan Mountains Region (HMR) uplift to environmental changes in the HMR and

its eastern adjacent area: Tracing the evolutionary history of *Allium* section *Sikkimensia* (Amaryllidaceae). *Molecular Phylogenetics and Evolution.* 130, pp. 380–396.

Xie, D.-F., Yu, H.-X., Price, M., Xie, C., Deng, Y.-Q., Chen, J.-P., Yu, Y., Zhou, S.-D., and He, X.-J., 2019. Phylogeny of Chinese *Allium* Species in section *Daghestanica* and Adaptive Evolution of *Allium* (Amaryllidaceae, Allioideae) Species Revealed by the Chloroplast Complete Genome. *Frontiers in Plant Science.* 10, pp. 1–15.

Yagi, M., 2018. Recent progress in whole genome sequencing, high-density linkage maps, and genomic databases of ornamental plants. 70, pp. 62–70.

Yagi, M., Kosugi, S., Hirakawa, H., Ohmiya, A., Tanase, K., Harada, T., Kishimoto, K., Nakayama, M., Ichimura, K., Onozaki, T., Yamaguchi, H., Sasaki, N., Miyahara, T., Nishizaki, Y., Ozeki, Y., Nakamura, N., Suzuki, T., Tanaka, Y., Sato, S., Shirasawa, K., Isobe, S., Miyamura, Y., Watanabe, A., Nakayama, S., Kishida, Y., Kohara, M., and Tabata, S., 2014. Sequence analysis of the genome of carnation (*Dianthus caryophyllus* L.). *DNA Research*. 21 (3), pp. 231–241.

Yakimowski, S.B. and Rieseberg, L.H., 2014. The role of homoploid hybridization in evolution: A century of studies synthesizing genetics and ecology. *American Journal of Botany*. 101 (8): pp. 1247-1258.

Yang, X., Zhang, R., Zhai, Z., Pang, Y., and Jin, Z., 2019. Machine learning for cultivar classification of apricots (*Prunus armeniaca* L.) based on shape features. *Scientia Horticulturae*. 256, p. 108524.

Ye, Y., Cai, M., Ju, Y., Jiao, Y., Feng, L., Pan, H., Cheng, T., and Zhang, Q., 2016. Identification and Validation of SNP Markers Linked to Dwarf Traits Using SLAF-Seq Technology in *Lagerstroemia*. *PLOS ONE*. 11 (7), p. e0158970.

Zalapa, J.E., Cuevas, H., Zhu, H., Steffan, S., Senalik, D., Zeldin, E., McCown, B., Harbut, R., and Simon, P., 2012. Using next-generation sequencing approaches to isolate simple sequence repeat (SSR) loci in the plant sciences. *American Journal of Botany.* 99 (2), pp. 193–208.

Zhao, M.L., Song, Y., Ni, J., Yao, X., Tan, Y.H., and Xu, Z.F., 2018. Comparative chloroplast genomics and phylogenetics of nine *Lindera* species (Lauraceae). *Scientific Reports*. 8 (1), pp. 1–11.

Zonneveld, B.J.M., 2008. The systematic value of nuclear DNA content for all species of *Narcissus* L. (Amaryllidaceae). *Plant Systematics and Evolution*. 275 (1–2), pp. 109–132.

Appendices

Appendix A

Three field expeditions to the *N. minor* group's distribution range were completed during this PhD project. In total 514 individuals were sampled from 47 populations under permit. Permits are provided electronically in Appendix A. Herbarium voucher material was submitted to the University of Reading herbarium (**RNG**). Details of the populations sampled are provided in Table A-1.

Table A-1 Details of sampled species and populations they were collected. Pop. =population code; N=number of individuals. * = species name which sample was collected.

Pop.	Species*	N	Location	Altitude	Date
AGA	*N. asturiensis*	3	Arange - Rio Mendo (43.23°, -8.01°)	261	11/03/2017
ALT	*N. asturiensis*	10	Alto Campoo (43.03°, -4.39°)	1912	19/05/2017
ARG	*N. asturiensis*	12	Monasterio de Leyre (42.65°, -1.18°)	1161	22/04/2016
CDC	*N. asturiensis*	10	Candanoso (43.05°, -4.89°)	1547	17/05/2017
CEB	*N. asturiensis*	10	Cebreiros (42.71°, -7.04°)	1300	18/03/2017
Entzia	*N. asturiensis*	15	Entzia (42.81°, -2.33°)	1027	16/04/2016
GOL	*N. asturiensis*	10	Golobar (42.97°, -4.35°)	1901	19/05/2017
ISI	*N. asturiensis*	10	Lagoa de Isoba (43.04°, -5.32°)	1376	16/05/2017
LUR	*N. asturiensis*	10	Luriana (43.08°, -4.83°)	1585	17/05/2017
PBL	*N. asturiensis*	10	Peña Blanca (42.83°, -4.80°)	1635	18/05/2017
PdF	*N. asturiensis*	15	Paso de Fraille (42.96°, -3.03°)	866	19/04/2016
PdH	*N. asturiensis*	10	Puerto de Herrera (42.60°, -2.68°)	1162	16/04/2016
PDM	*N. asturiensis*	10	Puerto de Monteviejo (42.91°, -4.92°)	1453	17/05/2017
PER	*N. asturiensis*	15	Sierra de Perdon (42.73°, -1.70°)	1006	16/04/2016
PMA	*N. asturiensis*	11	Piedramillera (42.64°, -2.33°)	824	20/04/2016
REG	*N. asturiensis*	10	Reguino - Rio Mendo (43.19°, -8.17°)	190	11/03/2017
ROZ	*N. asturiensis*	10	Camasobres (43.03°, -4.52°)	1331	18/05/2017
SAR	*N. asturiensis*	10	Sierra de Aralar (42.96°, -1.99°)	1015	16/04/2016

Table A-1 (cont.) Details of sampled species and the populations they were collected. Pop. =population code; N=number of individuals. * = species name which sample was collected.

Pop.	Species*	N	Location	Altitude	Date
SEN	*N. asturiensis*	10	Puerto de las Señales (43.08°, -5.24°)	1612	16/05/2017
SM	*N. asturiensis*	15	Sierra Salvada (43.02°, -3.18°)	809	19/04/2017
VM	*N. asturiensis*	15	Villamardones (42.87°, -3.24°)	859	19/04/2016
AGA	*N. cyclamineus*	10	Arange - Rio Mandeo (43.23°, -8.01°)	261	11/03/2017
BEM	*N. cyclamineus*	5	Bembibre - Rio Roubra (43.01°, -8.66°)	179	16/04/2017
CAN	*N. cyclamineus*	10	Codeseda - Rio Umea (42.63°, -8.44°)	385	17/03/2017
CUE	*N. cyclamineus*	12	Chelo - Rio Mandeo (43.27°, -8.17°)	13	11/03/2017
DEZ	*N. cyclamineus*	10	Laro - Rio Deza (42.62°, -8.21°)	387	17/03/2017
FRI	*N. cyclamineus*	14	Frixe - Rio Castro (43.00°, -9.24°)	28	16/03/2017
LER	*N. cyclamineus*	10	Acibeirero - Rio Lerez (42.62°, -8.31°)	703	17/03/2017
MON	*N. cyclamineus*	9	Mondariz - Rio Tea (42.24°, -8.45°)	34	17/03/2017
PON	*N. cyclamineus*	11	Ponte Couras – Rio Coras (41.88°, -8.70°)	90	13/03/2017
POV	*N. cyclamineus*	10	Povoa das Leivas (40.85°, -8.17°)	750	15/03/2017
SCO	*N. cyclamineus*	5	Serra de Ontes (42.88°, -8.94°)	120	16/03/2017
SDC	*N. cyclamineus*	3	Santa Catalina - Rio Xallas (43.05°, -8.84°)	326	16/03/2017
SIG	*N. cyclamineus*	11	Siguero - Rio Tambe (42.98°, -8.40°)	230	11/03/2017
SJM	*N. cyclamineus*	10	S. Joao Do Monte – Rio Agueda (40.60°, -8.24°)	552	15/03/2017
TEB	*N. cyclamineus*	7	Tebilhao (40.86°, -8.20°)	849	15/03/2017
XER	*N. cyclamineus*	1	Xerdiz (43.56°, -7.60°)	203	12/03/2017
BDL	*N. jacetanus*	15	Linza to Zuriza (42.89°, -0.80°)	1292	24/04/2016
BRA	*N. jacetanus*	10	Brañallin (42.98°, -5.78°)	1490	20/05/2017
PdLi	*N. jacetanus*	15	Puerto de Lizzarga (42.86°, -2.00°)	998	20/04/2016
PDP	*N. jacetanus*	10	Puerto de Piedrafita (43.03°, -5.61°)	1455	20/05/2017
SAN	*N. jacetanus*	10	Collado Santiago (42.91°, -5.48°)	1555	19/05/2017
VDA	*N. jacetanus*	15	Mtn Berrendi (42.97°, -1.22°)	1221	24/04/2016

Table A-1 (cont.) Details of sampled species and the populations they were collected. Pop. =population code; N=number of individuals. * = species name which sample was collected.

Pop.	Species*	N	Location	Altitude	Date
ARA	*N. minor*	10	Collada de Aralla (42.90°, -5.81°)	1553	20/05/2017
Aralar	*N. minor*	15	Sierra de Aralar (42.95°, -1.97°)	1174	16/04/2016
LDS	*N. minor*	10	Puerto de Escuredo (42.21°, -6.55°)	1829	21/05/2017
PAC	*N. minor*	10	Paerios (42.69°, -7.10°)	1071	18/03/2017
MDS	*N. minor*	11	Alto de las Berdiainas (42.37°, -6.46°)	1950	21/05/2017
RM	*N. minor*	4	Orbreiro - Rio Miño (43.03°, -7.61°)	360	12/03/2017
SDP	*N. minor*	10	Laguna de Los Peces (42.17°, -6.74°)	1720	21/05/2017

An additional 180 samples of wild origin material was sourced from collaborators for this project, outlined in Table A-2.

Table A-2 Details of sampled species and localities they were collected from collaborators. Pop. =population code; N=number of individuals.

Pop.	Species	N	Location	Collector
BD59	*N. abscissus*	1	Hort	B. Duncan
BD38	*N. alcaracensis*	1	Auvergne- France	B. Duncan
BD51	*N. alcaracensis*	1	Alcaraz- Spain	B. Duncan
BD05-06	*N. asturiensis*	2	Picos de Europa- Spain	B. Duncan
BD07-08	*N. asturiensis*	2	Alto de Faronpondo- Spain	B. Duncan
BD09, BD11-17	*N. asturiensis*	8	San Isidro- Spain	B. Duncan
BD10	*N. asturiensis*	1	Val de Ventrana- Spain	B. Duncan
BD18	*N. asturiensis*	1	Puerto de Puyo- Spain	B. Duncan
BD19, BD35	*N. asturiensis*	2	Monte Treito- Spain	B. Duncan
BD20	*N. asturiensis*	1	Manteigas- Portugal	B. Duncan
BD21-22	*N. asturiensis*	2	Gouveia- Portugal	B. Duncan
BD23	*N. asturiensis*	1	Torre from Velha- Portugal	B. Duncan

Table A-2 (cont.) Details of sampled species and localities they were collected from collaborators. Pop. =population code; N=number of individuals.

Pop.	Species	N	Location	Collector
BD46-48	*N. bicolor*	3	Puerto de Portalet- France	B. Duncan
BD49-50	*N. bicolor*	2	Astun- Spain	B. Duncan
BD36-37	*N. bujei*	2	Alcaraz- Spain	B. Duncan
BD43	*N. bujei*	1	Alcaraz- Spain	B. Duncan
BD27	*N. confusus*	1	Gredos- Spain	B. Duncan
BD53-55	*N. cyclamineus*	2	Portugal	B. Duncan
BD56	*N. cyclamineus*	1	Cangas- Spain	B. Duncan
BD57, BD60-61	*N. cyclamineus*	4	Galicia- Spain	B. Duncan
P2	*N. cyclamineus*	20	Calo (43.03, -9.14)- Spain	J. Loureiro
P3	*N. cyclamineus*	20	Bouza (43.14, -8.91)- Spain	J. Loureiro
P4	*N. cyclamineus*	20	Chaian (42.95, -8.53)- Spain	J. Loureiro
P5	*N. cyclamineus*	20	Caldas de Reis (42.60, -8.63)- Spain	J. Loureiro
P6	*N. cyclamineus*	20	Zamans (42.16, -8.69)- Spain	J. Loureiro
P7	*N. cyclamineus*	20	Porreiras (41.95, -8.56)- Portugal	J. Loureiro
BD24, BD29-32	*N. jacetanus*	5	Aisa- Spain	B. Duncan
BD25	*N. jacetanus*	1	Hort	B. Duncan
BD26	*N. lagoi*	1	Jaun- Spain	B. Duncan
BD28	*N. moleroi*	1	Col de la Cruet- France	B. Duncan
BD42	*N. nobilis*	1	Vieira Do Minho- Portugal	B. Duncan
BD44	*N. nobilis*	1	Bariviedo de los Reina- Spain	B. Duncan
BD45	*N. nobilis*	1	Tarnia- Spain	B. Duncan
BD40	*N. perez-chiscanoi*	1	Fuente del Agua- Spain	B. Duncan
RHS2	*N. poeticus* var. *hellenicus*	1	RHS Wisley; W2007-3140-A	K. Könyves

Table A-2 (cont.) Details of sampled species and localities they were collected from collaborators. Pop. =population code; N=number of individuals.

Pop.	Species	N	Location	Collector
RHS3	*N. poeticus* var. *physaloides*	1	RHS Wisley; W2008-2757-A	K. Könyves
BD33	*N. primigenius*	1	Senales- Spain	B. Duncan
BD34	*N. primigenius*	1	Puerto de Portillinos- Spain	B. Duncan
BD52	*N. primigenius*	1	Hort	B. Duncan
BD62-63	*N. provincialis*	2	Grasse- France	B. Duncan
BD39	*N. segurensis*	1	Hornos- Spain	B. Duncan
BD41	*N. varduliensis*	1	Vitoria- Spain	B. Duncan

In order to provide context to the study group, published *matK* sequences (Könyves, David, et al., 2019) were compiled with data gathered in this thesis. Furthermore, several samples were gathered during field expeditions during this project to use as outgroups. The details of the species sampled, and the voucher material are detailed in Table A-3.

Table A-3 Details of sampled species and localities they were collected from Könyves, David, et al. (2019).

Section	Species*	Location	Voucher
Angustifolii	*N. elegans*	NNE of Souk-El-Arbaa-du-Rharb - Morocco	RNG
Apodanthi	*N. cuatrecasasii*	VIllaluenga del Rosario - Spain	RNG
	N. rupicola	Poço do Inferno – Spain	K
	N. scaberulus	Girabolhos - Portugal	K
Aurelia	*N. brousonetii*	Commercial Source – Miniature Bulbs	RNG
Braxireon	*N. cavanillesii*	Taforalt to Grotte du Chameau - Morocco	RNG
Bulbocodii	*N. bulbocodium*	Mendiga- Portugal	RNG
	N. bulbocodium	Nuevo Portil- Spain	RNG
	N. bulbocodium	Aldea del Cano- Spain	RNG
	N. bulbocodium	Alconera- Spain	RNG

Table A-3 (cont.) Details of sampled species and localities they were collected from Könyves, David, et al. (2019).

Section	Species*	Location	Voucher
Bulbocodii (continued)	*N. bulbocodium*	Embalse de Cornalvo- Spain	RNG
	N. bulbocodium	Nogueira	RNG
	N. bulbocodium	Ourika valley, Amassine stream	RNG
	N. cantabriucus	Tiqqi to Immouzzèr-des-Ida-Outanane	RNG
	N. bulbocodium	Serra do Oribio- Spain	RNG
	N. bulbocodium	Near Mouronho- Portugal	RNG
	N. bulbocodium	Lagoa Comprida- Portugal	RNG
	N. bulbocodium	Catoira- Spain	RNG
	N. bulbocodium	Puerto de la Morcuera- Spain	RNG
	N. bulbocodium	Paso de Fraille- Spain	RNG
	N. cantabricus	Almogía to Villanueva de la Concepción- Spain	RNG
	N. cantabricus	Tighmi to Tafraoute- Morocco	RNG
	N. hedraeanthus	Quesada, Puerto de Tiscar- Spain	RNG
	N. romieuxii	Âïn Leuh- Morocco	RNG
	N. romieuxii	Ketama to Bab Berred- Morocco	RNG
	N. romieuxii	Khénifra to Aguelmame Azigza- Morocco	RNG
Dubii	*N. tortifolius*	Grazalema - Spain	RN
Ganymedes	*N. triandrus*	Sierra de la Garza - Spain	K
Jonquillae	*N. assoanus*	Ermita Virgen de la Sierra - Spain	RNG
	N. assoanus	Sierra de Perdon - Spain	RNG
	N. jonquilla	Rio Ayuela - Spain	K
	N. viridiflorus	c. 45km NNE of Kenitra - Morocco	RNG
Pseudonarcissi	*N. alpestris*	S. Juan de la Peña - Spain	RNG
	N. asturiensis	Lagoa Comprida - Portugal	K
	N. asturiensis	Alto de Cebreiro - Spain	K

Table A-3 (cont.) Details of sampled species and localities they were collected from Könyves, David, et al. (2019).

Section	Species*	Location	Voucher
Pseudonarcissi (continued)	*N. bicolor*	Sierra de Urbasa - Spain	RNG
	N. cyclamineus	Nogueira - Portugal	RNG
	N. cf. *pseudonarcissus*	Reserva de insectos - Spain	K
Tazettae	*N. papyraceus*	Villanueva de la Concepción - Spain	RNG
Intersectional hybrids	*N.* × *alleniae*	Not known	RNG
	N. × *intermedius*	Commercial source - Broadleigh Bulbs	RNG
	N. × *fosteri*	Puebla de Don Rodrigo to Arroba de los Montes- Spain	RNG

Appendices B-E are provided electronically.