Phylogenetics of paclitaxel biosynthesis genes in *Taxus baccata*, *Taxus* hybrids and allies.

Patricia Coughlan

Degree of M.Sc. (by Research)

University of Dublin, Trinity College

Supervisors: Prof. Trevor R. Hodkinson B.Sc., M.Sc., Ph.D.

Department of Botany, Trinity College Dublin

and

Dr. James Carolan B.Sc., Ph.D.

Department of Biology, Maynooth University

Submitted to the University of Dublin, Trinity College, June 2020

Declaration

I declare that this thesis has not been submitted as an exercise for a degree to this or any other university. It is entirely my own work except where indicated and clearly acknowledged in the text.

I agree to deposit this thesis in the University's open access intuitional repository or allow the library to do so on my behalf, subject to Irish Copyright Legislation and Trinity College Library conditions of use and acknowledgement.

I consent to the examiner retaining a copy of the thesis beyond the examining period, should they wish (EU GDPR May 2018).

_____

Patricia Coughlan

Acknowledgements

Firstly, I would like to express my sincere gratitude to my supervisors Dr. Trevor Hodkinson and Dr. James Carolan for their continuous and unwavering support. Thank you both for your patience, motivation, guidance and immense knowledge. I could not have imagined having better supervisors. I feel very fortunate to have worked with you both.

I would like to sincerely thank Ms. Ingrid Hook for provision of plant material and for general guidance and advice.

I would like to thank all the postgrads and staff of the Department of Botany, Trinity College Dublin, for your continued support and encouragement over the years. A special thank you to Dr. Daniel Kelly, Siobhan McNamee, Jackie Stone and Mark Kavanagh.

Thanks to the National Botanic Gardens, Dublin for allowing me to collect samples and to Dr. Noleen Smyth for showing me around.

A special thank you goes to my husband Trevor Hughes, my daughter Ava Hughes, my Mam & Dad and my family for their constant support and words of encouragement. Thank you to Roxy for always being there.

This thesis is dedicated to the memory of Maximillion Von Sternberg, Dr. Caoimhe Muldoon, Dougie Hughes and Lily Hughes who always believed in my ability to be successful in my research. You all are gone but your belief in me has made this journey possible.

Abstract

Taxus L. is a genus of trees and shrubs with high value in horticulture, and in medicine as a source of the anticancer drug Paclitaxel. The taxonomy of the group is highly complex due to the lack of diagnostic morphological characters and the high degree of phenotypic plasticity among species. This often leads to misidentification and problems with classification using traditional taxonomic methods that rely solely on morphological characters. The complexity has become an issue for the pharmaceutical industry because of the anti-cancer properties of Paclitaxel. Paclitaxel is a diterpenoid, which is produced by several species of *Taxus* and is used to treat various forms of cancer. *Taxus* has a wide global geographic distribution and some taxonomists recognise only a single species with geographically defined subgroups. However, others have preferred to split the genus into several species. To address these differences, we conducted a thorough phylogenetic analysis (Maximum Likelihood, Bayesian Inference and TCS haplotype network analyses) of all its species and many subtaxa using nuclear and plastid gene regions (nr*ITS* and plastid *trnL* intron and *trnL-F* intergenic spacer). Results support the recognition of nine distinct species (*T. baccata, T. brevifolia*, *T. canadensis*, *T. cuspidata*, *T. floridana*, *T. fauna*, *T. globosa, T. sumatrana* and *T. wallichiana*) but evidence is found for less species distinction and considerable reticulation within the T. *baccata*, *T. canadensis* and *T. cuspidata* group. There is some biogeographic structure in the nr*ITS* data showing that *Taxus brevifolia* is sister to *T. globosa* and *T. floridana. Taxus fauna* groups with *T. contorta. Taxus wallichiana* is resolved as monophyletic but its varieties *T. wallichiana* var. *mairei,* var. *chinensis* and var. w*allichiana* are not monophyletic although individuals within varieties generally group well together. Evidence is also presented for the sister group status of *Pseudotaxus* to *Taxus* and the inclusion of *Amentotaxus*, *Austrotaxus*, *Cephalotaxus* and *Torreya* within Taxaceae. We compare the results to known taxonomy, and present preliminary new leaf anatomical data using leaf impressions to visualise epidermal and stomatal characters. We also investigate the hybrids *T.* x*media* and T. x*hunnewelliana* and present nr*ITS* data identifying the origin of these taxa and their putative parental species.

The biosynthesis of Paclitaxel involves 19 steps. We characterised two genes, Taxadiene synthase (*TS*) and 10-deacetyl baccatin III-10-0-acetyltranferase (*DBAT*), involved in the pathway and compared variation in the genes among species of *Taxus. TS* is involved in the first committed step of the pathway. We developed new molecular PCR primers to amplify and study the *TS* 1 to 5 exon regions. The primary aim was to assess molecular variation in the *TS* gene at DNA and protein levels and to test for evidence of selection on the gene. However, we also sequenced a section of the *DBAT* gene, for a sample of taxa, to look for evidence of molecular variation in that gene and its presence in relatives of *Taxus*. The substitution rate is high in the *TS* gene with on average 43 changes per kbp and there is considerable variation in the amino acid translation, which varies with exon. *TS* gene trees based on all codon positions are not consistent with the *ITS* and *trnL-F* trees hinting at selection acting on the gene. Our results show that there is much variation in the sequences of *TS* and *DBAT* across species, that many are non-synonymous and that these changes are possibly due to parallel changes caused by selection on the genes.

These results bring more clarification to the taxonomy of *Taxus* and highlight the high levels of variation in Paclitaxel genes. It may provide the pharmaceutical industry with promising targets for genetically engineering more efficient biosynthetic production of Paclitaxel and its precursors.

Table of Contents

List of Figures

List of Tables

Chapter 1: General Introduction

1.1. Introduction

Taxus baccata Linnaeus is more commonly known as the yew tree and is a slow growing conifer sometimes referred to, in folklore, as the "tree of eternity" due to its longevity. It has been associated with birth, death and immortality (Hageneder, 2007). It is commonly found growing in church yards and has been important to humans for thousands of years. In ancient Ireland, yew was used to carve household items such as bowls and spoons (Hageneder, 2005) as well as weapons for hunting. One of the oldest wooden artefacts in the world is a hunting weapon made of yew, The Clacton Spear, approximately 400,000 years old (Allington-Jones, 2015). The yew tree is traditionally seen as a plant with paradoxical properties as it was viewed to represent both death and rebirth, provide poison and medicine, weapon and shield (MacCoitir, 2003). All parts of the yew tree are poisonous except for the fleshy aril and yet it is what makes the yew poisonous which makes the yew tree so important to us in modern times, as it is used to treat cancer (Wani et al., 1971, Rowinsky et al., 1990).

On the 21st of August 1962, Dr. Arthur Barclay gathered samples of needles, stems and bark from a Pacific yew within the Gifford Pinchot National Forest, Washington, USA. It was part of the American government's programme to find cures for cancer from wild species. Dr. Monroe Wall extracted a compound he named Taxol (Paclitaxel) from the bark of Pacific yew (*T. brevifolia* Nutt.) samples and it was subsequently found to be a cancer killing agent (Wani et al., 1971). From 1960 to 1981 the American government's programme screened over 114,000 plant extracts and 16,000 animal extracts with Taxol being the primary compound of interest after 20 years work. Bristol Myers Squibb further developed Taxol commercially and gave it the trade-mark name Taxol. The generic name of Taxol was then changed to Paclitaxel. Taxol is used to treat ovarian, breast and lung cancer (McGuire et al., 1989, Holmes et al., 1991, Bristol Myers Squibb, 2011).

Bristol Myers Squibb developed a method to extract Paclitaxel from the needles of *T. baccata* L. (Denis et al., 1988, Denis et al., 1990) instead of the bark of *T. brevifolia*, as approximately 3-6 mature trees of *T. brevifolia* were required to obtain 1 gram of Taxol needed to treat one patient (Cragg et al., 1993). It was taking 30 ton shipments of bark

to produce up to 100 grams of purified Taxol, which was causing the tree to become endangered and was clearly not a sustainable option for its production (Cragg and Snader, 1991, Goodman and Walsh, 2001).

Taxus baccata is also a natural producer of Paclitaxel (Witherup et al., 1990, ElSohly et al., 1997, van Rozendaal et al., 2000). The first phytochemical study on the foliage of *Taxus baccata* was carried out by Lucas (1856). He isolated an alkaloid mixture which he called taxine. Taxine is a mixture of fractions including Paclitaxel, taxine A and taxine B (Graf and Boeddeker, 1956) as referred to in Wilson et al. (2001). Taxine B is much more cardiotoxic than taxine A (Alloatti et al., 1996). The cardiotoxic alkaloid taxine B is abundant in *T. baccata* but is almost absent in *T. brevifolia* (Tyler, 1960) as referenced in Itokawa and Lee (2003). If *T. baccata* had been sampled instead of *T. brevifolia* it might have been thought to be too toxic and Taxol might never have succeeded in becoming a cure for cancer. Paclitaxel is found in all parts of the yew tree except for the fleshy aril. If it is ingested it can cause death by heart failure. Paclitaxel has been shown to be embryotoxic and fetotoxic in rabbits, and to decrease fertility in rats (Irish Medicines Board, 2007).

Some knowledge has been generated about the phylogenetic relationships of *Taxus* species (Kilmartin, 2002, Hao et al., 2008a), however very little is known about molecular DNA variation in the genes leading to Paclitaxel production in the genus. This project adds to the existing knowledge of Paclitaxel by developing molecular primers to amplify and study the genes involved in the Taxol biosynthetic pathway (Figure 1.1). It took a phylogenetic approach to study genetic variation in these genes that might be important to variation in Paclitaxel production. More specifically, it assessed variation in these genes between *T. baccata* and *Taxus* hybrids such as *T. x media* Rehder. The focus of the study was on the molecular variation between the genes involved in the production of Paclitaxel in *T. baccata*.

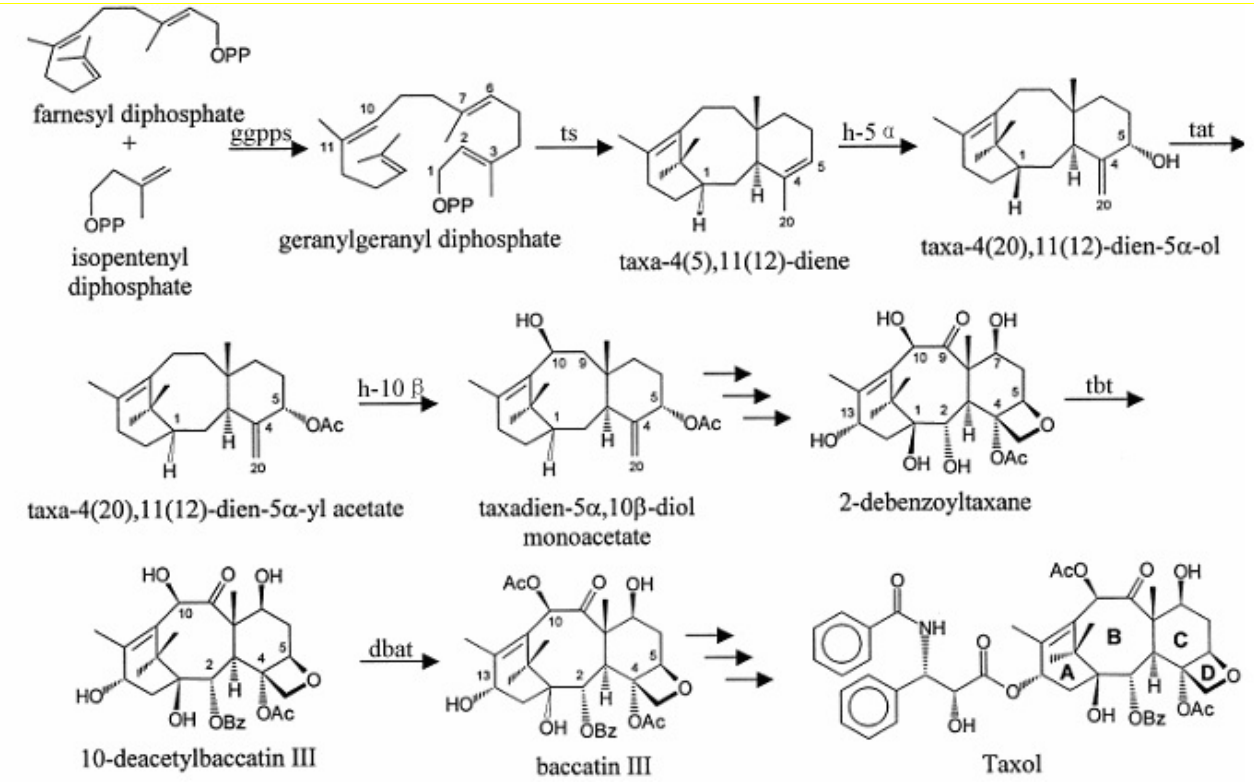


Figure 1.1. Taxol biosynthetic pathway and its genes. ggpps: geranylgeranyl diphosphate synthase; ts: taxadiene synthase; h-5_: cytochrome P450 taxadiene 5_-hydroxylase; tat: taxa-4(20), 11(12)-dien-5a-ol-O-acetyltransferase; h-10_: cytochrome P450 taxane 10_-hydroxylase; tbt: taxane 2a-O-benzoyltransferase; dbat: 10-deacetyl baccatin III-10-O-acetyltransferase. Multiple arrows indicate several as yet undefined steps (Guo et al., 2006). Further details of the Taxol biosynthesis pathway can be found in Walker and Croteau (2001), Jennewein et al. (2004), Croteau et al. (2006), Nims et al. (2006) and Guo et al. (2006).

This project follows on from the MSc work of Kilmartin (2002) who studied the phylogenetics of *Taxus baccata*, under the supervision of Drs T. Hodkinson and I. Hook, using non-coding DNA sequences that determined the relationships of species within the genus. That MSc did not fully resolve phylogenetic pattern in the genus nor did it study the Paclitaxel genes of *Taxus* because markers were not available at the time to do so.

Variation in these genes can now be studied because recent advances in molecular biology have elucidated the taxol biosynthetic pathway (Figure 1.1) and many of its genes have been characterised in a limited number of species (Wildung et al., 1996, Hefner at al., 1998, Walker et al., 2000). This thesis aimed to study molecular variation in Paclitaxel genes to establish differences between species, including hybrids and their known parents, so that associations can be discovered between Paclitaxel production and DNA sequence polymorphism. The primary aim was to discover variation in paclitaxel genes in *T. baccata* and its close relatives. Two candidate genes known as taxadiene synthase (*TS*) and 10-deacetylbaccatin III-10b-O-acetyltransferase (*DBAT*) have been assessed in a number of species (Figure 1.2). We also aimed to examine both inter-specific and intra-specific variation and the inheritance of polymorphisms in hybrids such as *Taxus* x *media*. Two other gene regions, namely the nuclear *ITS* (internal transcribed spacer region of nrDNA) region and the plastid *trnL-F* (a transfer RNA gene region) were also studied to help with species identification and phylogenetic reconstruction so that the evolution of *TS* could be interpreted in a phylogenetic context. This project had partial success in sequencing *DBAT* in *Taxus* but mainly concentrates on the *TS* gene because it was the only region which sequenced consistently across a broad range of yew taxa.

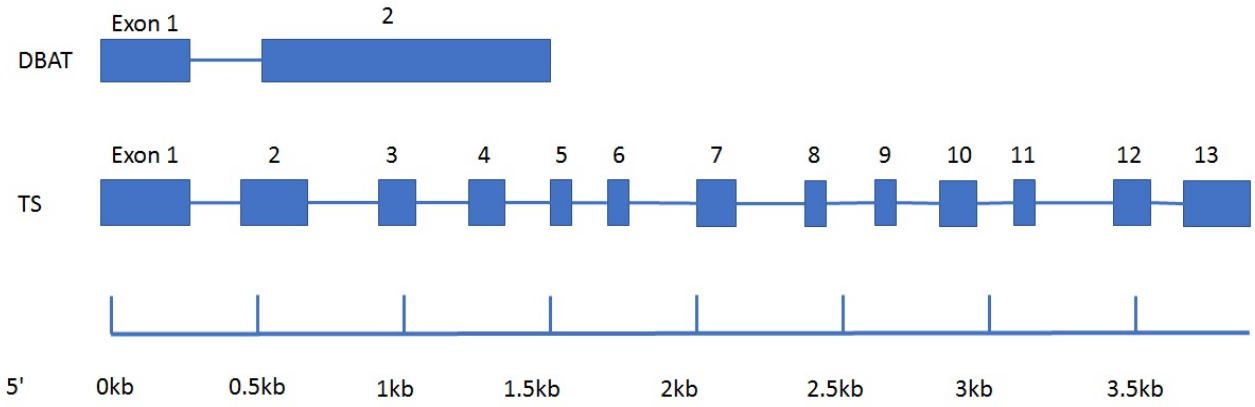


Figure 1.2. Exon/intron structure of TS and DBAT genes in Taxus drawn to scale. Boxes indicate exons; areas amplified for interspecific analysis are exons 1 and 2 of DBAT and exons 1–9 of TS. Adapted from (Hao et al., 2009).

1.2. **Taxonomy**

Taxus baccata belongs to the Taxaceae family (Farjon, 1998). Taxaceae is a family of evergreen shrubs or trees, which are mainly dioecious. The bark is thin and exfoliates in strips. The seed is contained in a fleshy aril which is red, purple or yellow when ripe (Farjon, 2017). According to Eckenwalder (2009) there are six genera of Taxaceae *Amentotaxus, Cephalotaxus, Pseudotaxus, Taxus, Austrotaxus* and *Torreya.* Farjon (2017) has five genera in Taxaceae. He classifies *Cephalotaxus* in a family of its own. Pilger (1903) includes *Cephalotaxus* in Taxaceae, however Pilger (1926) assigned it a separate family of its own.

Taxus is composed of several closely related species found growing naturally across the northern temperate region of the globe (Cope, 1998). Species include the European yew, *T. baccata,* and the pacific yew, *T. brevifolia. Taxus baccata* is native to Europe, western Asia and North Africa. *Taxus brevifolia* is native to North America. The total number of species and varieties of *Taxus* is unclear. According to the International Plant Name Index (www.ipni.org; November 2016) there are 195 different named *Taxus* but this list includes all names and also synonyms. The number of recognised species varies considerably depending on author. For example, according to Spjut (2007b) *Taxus* has 24 species and 55 varieties. Farjon (1998), Farjon (2017) recognises *Taxus* as having 10

species. The Plant List (http://www.theplantlist.org) lists 73 species names but accepts just nine of these as listed below and shown in Figure 1.3:

Taxus baccata L.

Taxus brevifolia Nutt.

Taxus canadensis Marshall

Taxus cuspidata Siebold & Zucc.

Taxus floridana Nutt. ex Chapm.

Taxus fuana Nan Li & R.R.Mill

Taxus sumatrana (Miq.) de Laub.

Taxus globosa Schltdl.

Taxus wallichiana Zucc.

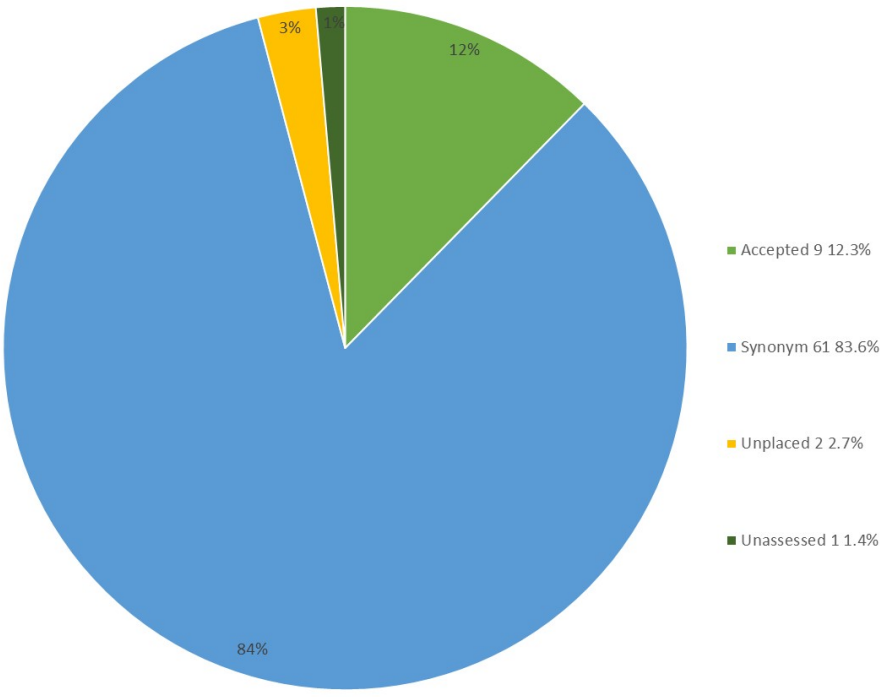


Figure 1.3. Described Taxus species, accepted species and synonyms (Source: The Plant List http://www.theplantlist.org)

However, there is dispute as to whether *Taxus* has far fewer species. Some authors argue for the recognition of distinct geographical forms of several species instead of recognising these at specific rank (Pilger, 1903). For example, *T. baccata* is mainly European, *T. brevifolia* is found in pacific North America and *T. canadensis* in eastern North America. See in Figure 3.1, page 54.

Taxus has high levels of phenotypic plasticity (Collins et al., 2003). This means that *Taxus* morphology can vary considerably depending on its environment. This often leads to misidentification by traditional taxonomic methods relying on morphological characters. Delimitation of species in *Taxus* has been a long standing taxonomic problem. Pilger (1903) classified *Taxus* as monotypic, a single species with seven geographical subspecies. Elwes and Henry (1906) were also of the opinion that the genus was comprised of a single species with distinct geographical forms. Dempsey and Hook (2000) conformed to this opinion. Species delimitation is a major problem as it is holding back progress in pharmaceutical studies. Appendino (1995) and Dempsey and Hook (2000) highlighted this problem in the context of drug development.

Some molecular work has been done to help correctly identify *Taxus* species. Chaw et al. (1993) and Cheng et al. (2000) did molecular work which clarified the position of Taxaceae as it relates to other conifers and placed it in the order Taxales in the gymnosperms. Li et al. (2001) did some molecular phylogenetic analyses using plastid *trnL-F* sequences and showed genetic diversity among species. Dempsey (2000) conducted a study to look at the morphological and phytochemical characteristics of 33 species and varieties of *Taxus* to see whether they could be clearly distinguished from each other and from putative sister genera *Cephalotaxus* and *Torreya.* He found no single parameter of taxonomic value which enabled the conclusive distinction between one *Taxus* species and another. However, he did find that microscopic and phytochemical characteristics could be used to distinguish between *Taxus* and other closely related genera (Dempsey, 2000).

Collins et al. (2003) conducted a study that looked at the species separation between *Taxus baccata*, *T. canadensis*, and *T. cuspidata* and their reputed hybrids and attempted to find a reliable molecular method to correctly identify *Taxus* species, hybrids and cultivars. A total of 19 samples from three species (*T. baccata*, *T. canadensis*, and *T. cuspidata*) and 31 samples from putative hybrids (*T.* x *hunnewelliana* and *T.* x *media*) were DNA fingerprinted using random amplified polymorphic DNA fingerprinting (RAPDs) and sequenced for the plastid *trnL-F* gene region. All samples showed unique

RAPD banding profiles. 21 RAPD bands were species specific. Hybrid origins and parentage were able to be confirmed by the presence of these bands. They reported that 10 different primers resulted in a data set of 201 individual bands of which 185 were polymorphic. The lowest variation within species was among individuals of *Taxus canadensis*. The highest variation was seen in *Taxus baccata.* They reported a close association between *T. canadensis* and *T. cuspidata* with both species equidistant to *T. baccata.* In the chloroplast *trnL-F* region, 16 polymorphic sites were found between species and three different haplotypes were identified.

1.3. Genus and species description

Taxus baccata Linnaeus

Taxus baccata is an evergreen shrub or small tree and may have bark peeling off in scales (Figure 1.5). It is 20-29m tall and its trunk is 1.5-3.5m in diameter. It is usually dioecious, having separate male and female plants. Its leaves are flat, soft, dark green and approximately 1 to 4cm long. The male cones are small and globular with approximately 10 pollen-producing organs (microsporophylls). The female reproductive structures consist of a single ovule with small scales at the base. The green seed is surrounded by a red aril. Its crown is broad, dense and rounded. Its needle like leaves are flat, soft and alternate. They are arranged in two flanks in a flattened spray (Eckenwalder, 2009, Parnell and Curtis, 2012).

Figure 1.4. Lithograph print of Taxus baccata (Krausel, 1960)

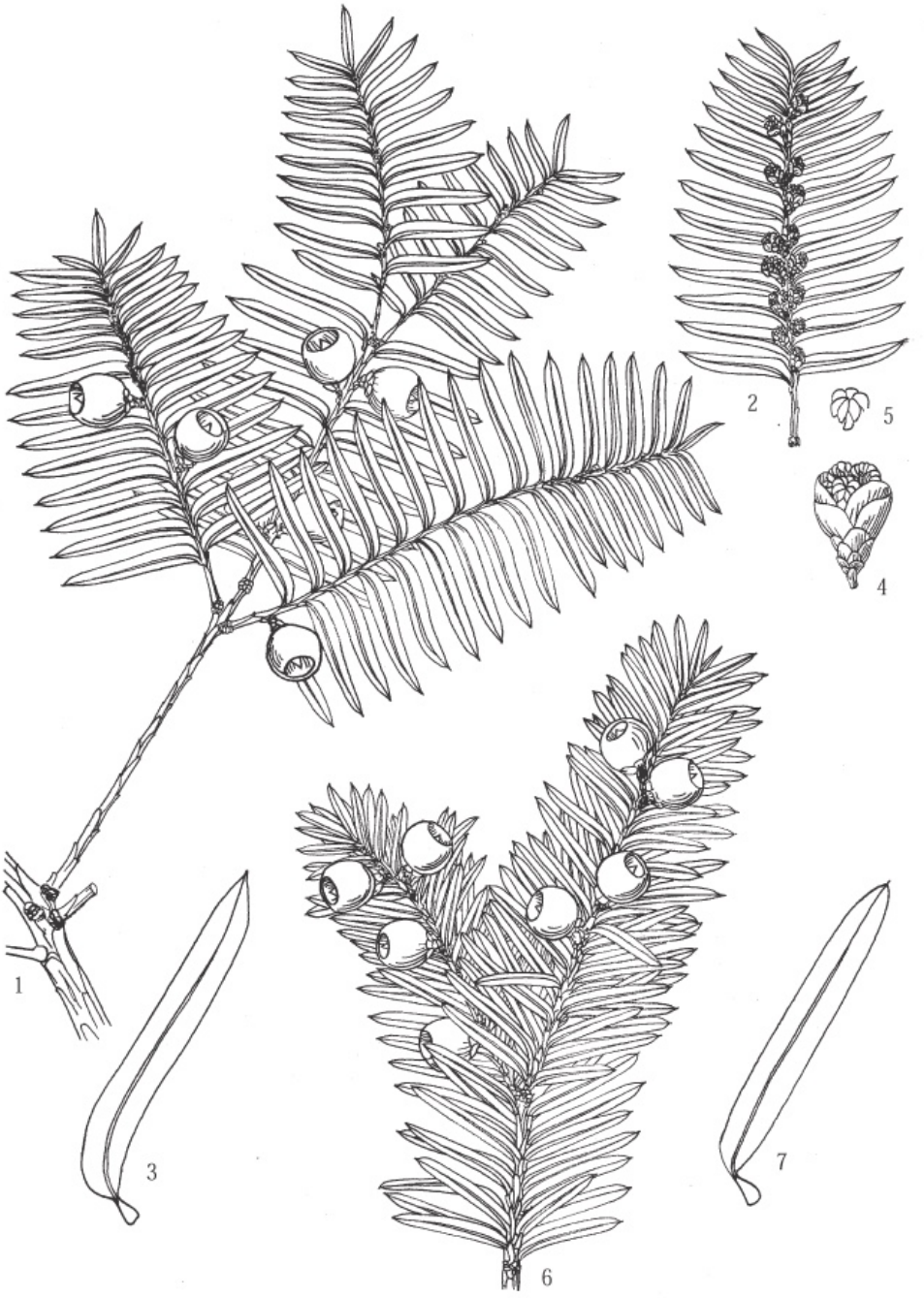


Figure 1.5. Taxus morphology. 1-5: T. wallichiana (1,3 seed/aril bearing branchlets and leaf); 2,4,5 cone bearing branchlet, pollen cones, microsporophyll; 6-7 T. cuspidata (seed/aril bearing branch, leaf). (Fu et al., 1999)

Source: Flora of China http://www.efloras.org/object_page.aspx?object_id=1351&flora_id=2

Taxus baccata cultivars

According to the World Conifer Data Pool 1993 there are more than 200 described forms of *Taxus baccata* (Welch et al., 1993) including, for example, the following commonly used cultivars:

'Lutea' – Yellow berries

'Aurea' – The golden yew has yellow needles

'Fastigiata' – The Irish yew which grows very upright, forming wide columns. The first Irish 'Fastigiata' yew was discovered in County Fermanagh in 1780 (Nelson, 1993)

Taxus brevifolia T.Nuttall

Taxus brevifolia is a dioecious shrub or tree to 15-25m tall with a trunk 0.6 -1.5m in diameter. The outer bark is scaly and purplish to brown in colour while the inner bark is reddish to reddish purple. The seed is clearly visible within an orange or red (rarely yellow) aril. Its crown is conical to dome-shaped becoming open and irregular with age. The needles are shorter than those of *T. baccata* (Farjon, 2017) hence the name *brevifolia* which is Latin for with short leaves (Eckenwalder, 2009).

Taxus canadensis H.Marshall

The Canadian yew is a low sprawling monoecious shrub, 1-2m tall. This species is distinguished from all other yews by its lack of a trunk, lack of microscopic papillae on the cells of its stomata bands and by having pollen cones and seeds on the same plant. Its crown is low and wide with thin spreading branches turned up at the end. Its seeds are clearly visible within a bright red aril (Eckenwalder, 2009).

T. cuspidata Siebold & Zucc

The Japanese yew is a dioecious shrub or tree (Figure 1.6). It grows up to 20m tall and has a trunk 1m in diameter. Its crown is deep and narrow in forest grown trees but can be dense to open and irregular, rounded or flat-topped. Its needle like leaves are in a V-shaped position. The seeds are clearly visible within a bright red or yellow aril (Eckenwalder, 2009). Its name refers to the leaves terminating in a small cusp (Farjon, 2017).

Taxus x hunnewelliana Rehder

Is believed to be a hybrid between the Canadian yew *Taxus canadensis* and *T. cuspidata* the Japanese yew. The original cross was made at Hunnewell Pinetum Massachusetts and described by Alfred Rehder in 1925 (Rehder, 1925).

Taxus x media Rehder

Is believed to be a hybrid between the European yew *Taxus baccata* and the Japanese yew *T. cuspidata.* It was developed by T.D Hatfield in Wellesley Massachusetts in the early 1900s (Rehder, 1923).

Taxus wallichiana Zuccarini

The Himalayan yew is a dioecious shrub or a tree 2-30m tall (Figure 1.6). Its trunk is 1.5-3.5 m in diameter. It has a dome shaped dense crown with thin upwardly angled branches. The branchlets are horizontal or drooping. The seeds are clearly visible within a red aril. A synonym for *Taxus wallichiana* is *Taxus fuana* (Eckenwalder, 2009) but is considered a separate species in this thesis and in the Plant List.

The Flora of China divides *T. wallichiana* into three varieties namely:

T. wallichiana var. *wallichiana* Zucc. Distributed in SW Sichuan, SE Xizang, W Yunnan, Bhutan, N India, N Myanmar, Sikkim, S Vietnam

T. wallichiana var. *mairei* (Lemée & H. Léveillé) L. K. Fu & Nan Li. Distributed in S Anhui, Fujian, S Gansu, N Guangdong, N Guangxi, Guizhou, W Henan, W Hubei, Hunan, Jiangxi, S Shaanxi, Sichuan, Taiwan, E Yunnan, Zhejiang [N India, Laos, Myanmar, Vietnam].

T. wallichiana var. *chinensis* (Pilger) Florin. Distributed in S Anhui (Huang Shan), Fujian, S Gansu, N Guangxi, SE and W Guizhou, W Hubei, NE Hunan, S Shaanxi, Sichuan, E Yunnan, Zhejiang; cultivated in Jiangxi (Lu Shan) [N Vietnam].

Taxus globosa D.F.L Schlechtendal

The Mexican yew is a dioecious shrub or tree 20-40m tall with a trunk 0.4m in diameter. It has a narrow and irregular cylindrical crown. Its branches are slender and horizontal. Its seeds are clearly visible within a bright red aril (Eckenwalder, 2009).

Taxus floridana A.W. Chapman

Taxus floridana also known as the Florida yew is one of the rarest conifer species. It is a dioecious shrub or tree 6-10m tall with a trunk of 0.2-0.4m in diameter. Its seeds are clearly visible within a dull red aril. It has a narrow dome shaped crown which is open and irregular. Its branches are slender and angled upwardly. Its branches are horizontal (Eckenwalder, 2009).

Taxus sumatrana (Miquel) de Laubenfels

The Chinese yew is a shrub or tree 30-45m tall. Its trunk is 1-2m in diameter. It has a dense dome shaped crown which is upwardly angled. Its branches are horizontal. It is dioecious. Its seeds are clearly visible within a red aril (Eckenwalder, 2009).

Pseudotaxus W.C Cheng

The white – cup yew is an evergreen shrub or tree. It is a dioecious plant. It has several slender trunks which can be contorted and branch repeatedly from near the base. Its crown is dense and dome shaped. It has many thin branches bearing single, paired or clustered branchlets. Its seed is surrounded by a fleshly white cup-shaped aril (Eckenwalder, 2009).

Cephalotaxus P.Siebold & Zuccarini ex Endlicher

Cephalotaxus plants are dioecious. They are evergreen trees and shrubs which have multiple stems near the ground. It is the only genus in the Taxaceae with seed cones. There are five species of *Cephalotaxus,* C. *oliveri,* C. *sinensis,* C. *mannii,* C. *harringtonii* and C. *fortunei* (Eckenwalder, 2009).

Austrotaxus R.H.Compton

Austrotaxus also known as New Caledonian yew, gets its scientific name from Latin meaning southern yew. It is the only genus of Taxaceae native in the southern hemisphere. It is a dioecious evergreen tree or shrub. It grows 15-25m tall and has a trunk 0.3-0.7m in diameter. The seeds are enclosed in a deep purple aril (Eckenwalder, 2009).

Podocarpus L'Heritier ex Persoon

Podocarpus are dioecious evergreen shrubs and trees. There are 82 species of *Podocarpus* (Eckenwalder, 2009).

Torreya G.Arnott

The nutmeg yew is dioecious or rarely monoecious evergreen trees or shrub. It has one to few cylindrical to off-center trunks (Eckenwalder, 2009). The seed is completely surrounded in a fleshy purplish to bluish black aril (Farjon, 2017). There are 6 species of *Torreya* according to the Plant List *Torreya californica* Torr, *Torreya grandis* Fortune ex Lindl, *Torreya jackii* Chun , *Torreya nucifera* (L.) Siebold & Zucc. , *Torreya parvifolia* T.P.Yi, Lin Yang & T.L.Long and *Torreya taxifolia* Arn.

Amentotaxus Pilger

The catkin yew. A dioecious evergreen small tree or shrub (Eckenwalder, 2009). The seed is completely hidden in a fleshy red or purple aril (Farjon, 2017). There are six species according to the Plant List *Amentotaxus argotaenia, A. assamica* D.K.Ferguson, *A.formosana* H.L.Li, *A.hatuyenensis* T.H.Nguyên, *A.poilanei* (Ferré & Rouane) D.K.Ferguson, and *A. yunnanensis* H.L.Li.

1.4. **Aims**

The focus of the study was therefore on the phylogeny of *Taxus* and allied genera and the molecular variation within the genes involved in the production of Paclitaxel in *T. baccata* (and on variation in these genes among species) so that associations can be discovered between Paclitaxel production and DNA polymorphism. More specifically the thesis aimed to:

- Develop markers to amplify and sequence genes involved in the biochemical pathway of paclitaxel production.
- Sequence *Taxus* taxa for plastid and nuclear genes (*trnL-F*, nr*ITS* respectively) and use these for phylogenetic reconstruction of the group and an assessment of biogeographical variation.

- Discover which existing Paclitaxel gene primers work in *Taxus* and outside the group e.g. *Cephalotaxus*, *Podocarpus, Torreya* and *Austrotaxus.*
- Align a matrix of the sequences for the *TS* and *DBAT* Paclitaxel genes and see how variable the genes are and how they have evolved.
- Examine variation in the Paclitaxel markers and see how different species differ from each other in terms of amino acid sequence. The variation can also be used for DNA barcoding and taxonomic purposes.

Chapter 2: Materials and Methods

2.1. Plant material

Fresh plant material of *Taxus* and related genera was collected from the National Botanic Gardens in Glasnevin, Dublin, Ireland in December 2011. Some DNA samples were also available from the Trinity College Dublin (TCD) DNA Bank. Ground and dried leaf samples were also provided by Ingrid Hook (TCD). Samples collected by Lisa Kilmartin, which had been stored in a -80^{o}C freezer, were also used. Most samples were *Taxus* but *Amentotaxus*, *Cephalotaxus*, *Podocarpus* and *Torreya* were included as outgroups. These are closely related to *Taxus* but not part of the ingroup and hence are useful for comparative purposes and rooting phylogenetic trees (Judd et al., 2016). *Cephalotaxus* is believed to be sister to *Taxus* (Hao et al., 2008b) and Taxaceae comprises *Amentotaxus, Taxus* and *Torreya* (Farjon, 1998). *Cephalotaxus* is sometimes included in Taxaceae (Li and Fu, 1997). *Podocarpus* belongs to Podocarpaceae (Farjon, 2010b). Voucher specimens were kept for each sample, dried and stored in the TCD Herbarium. A list of all samples used in this thesis is given in Table 2.1.

Table 2.1 List of samples used in thesis. Samples newly sequenced for ITS and trnL intron and trnL-F intergenic spacer or used in the morphological assessments

	Name	Voucher or living specimen, Accession ID	GenBank Code *ITS*; *trnL-F*
P1	*T. canadensis*	Glasnevin Hook *s.n.*	**MK116527**; MK731927
P2	*T. canadensis* 'Aurea'	Glasnevin Hook *s.n.*	**MK123469**; MK731928
P3	*T. cuspidata*	JFK Hook *s.n.*	**MK116528**
P5	*T. canadensis*	JFK Hook *s.n.*	**MK116529**; MK731929
P6	*T. canadensis*	Canada Hook *s.n.*	**MK116530**; MK731930
P8	*Cephalotaxus harringtoria* 'Fastigiata'	XX.006542 Glasnevin Coughlan P8	**MK116531**; MK731931
P9	*T. cuspidata f. thayerae*	1952.006540 Glasnevin Coughlan P9	**MK168608, MK168609**; MK731932
P10	*T. canadensis* 'Aurea'	XX.006549 Glasnevin Coughlan P10	**MK168610, MK168611, MK168612**; MK731933
P11	*T. canadensis*	XX.006556 Glasnevin Coughlan P11	**MK168613, MK168614, MK168615**; MK731934
P12	*T. cuspidata var. luteobaccata*	1930.006571Glasnevin Coughlan P12	**MK168616, MK168617, MK168618**; MK731935
P13	*T.* 'Aurea'	XX.006570 Glasnevin Coughlan P13	MK731936
P14	*T. brevifolia*	1885.006579 Glasnevin Coughlan P14	**MK123470**; MK731937
P15	*T. cuspidata*	XX.006597 Glasnevin Coughlan P15	**MK123471**; MK731938
P16	*T. x media* 'Hicksii'	XX.006589 Glasnevin Coughlan P16	**MK123472**; MK731939
P17	*T. cuspidata*	1911.006591 Glasnevin Coughlan P17	**MK123473**; MK731940
P18	*T. cuspidata f. thayerae*	1952.006593 Glasnevin Coughlan P18	**MK168781, MK168782, MK168783**; MK731941
P19	*T. cuspidata*	XX.006601 Glasnevin Coughlan P19	**MK168784, MK168785, MK168786**; MK731942
P20	*T. canadensis* 'Aurea'	XX.006609 Glasnevin Coughlan P20	MK731943
P21	*T. canadensis*	XX.005502 Glasnevin Coughlan P21	**MK21148, MK211149, MK211150**; MK731944
P22	*T. x media* 'Cuftoni'	XX.0055077 Glasnevin Coughlan P22	**MK211151**; MK731945
P23	*Podocarpus salignus*	XX.007572 Glasnevin Coughlan P23	MK731946
P24	*Torreya californica*	XX.007746 Glasnevin Coughlan P24	**MK211152**; MK731947
P25	*Unlabelled Torreya Sp.*	XX.007599 Glasnevin Coughlan P25	**MK211153**; MK731948
P26	*Podocarpus macrophyllus*	2005.0135 Glasnevin Coughlan P26	MK731953
P27	*T. baccata f.aurea*	XX.006574 Glasnevin Coughlan P27	**MK211154, MK211155, MK211156**; MK731949
P28	*T. baccata* 'Fastigiata'	13638 Glasnevin Coughlan P28	**MK211157**; MK731950
P29	*T. baccata* 'Grandis'	2001.4181 Glasnevin Coughlan P29	MK731951
P30	*T. baccata* 'Fastigiata'	Glasnevin Cemetery Coughlan P30	MK731952

L10	*T. baccata*	Ranelagh Park, Dublin	MK783697
L24	*T. canadensis*	Bedgebury 14/94	**MK272737**; MK748448
L25	*T. x hunnewelliana*	Bedgebury 13/125	MK783698
L26	*T. x media* 'Hicksii'	Bedgebury 13/006	MK783699
L27	*T. cuspidata*	Bedgebury 13/136	MK783700
L31	*T. brevifolia*	Glasnevin 1885.006579	MK783701
L32	*T. cuspidata* 'Fructo luteo'	Glasnevin 12/30Kew	MK783702
L33	*T. canadensis*	Glasnevin XX.006556	**MK272738**; MK748449
L34	*Torreya nucifera*	Glasnevin XX.007734	**MK272739**
L36	*T. canadensis* 'Aurea'	Glasnevin XX.006549	**MK272740**; MK783703
L42	*T. baccata* 'Fastigiata'	Bedgebury 13/108	MK783704
L40	*T. x media* 'Hatfieldii'	Bedgebury 13/280	_
L44	*T. baccata* 'Rushmoor'	Bedgebury 15/0287	MK783705
L51	*T. baccata*	Bedgebury 13/025	MK783706
L66	*Amentotaxus formosana*	Edinburgh 19763745	MK748450
L68	*Cephalotaxus sinensis*	Edinburgh 19081025	_
L73	*Torreya jackii*	Edinburgh 19970112	_
L75	*Cephalotaxus oliveri*	Edinburgh 951689	_
L93	*T. baccata*	Cornalack Lough Derg *s.n.*	MK748451
L94	*T.baccata*	Cornalack lough Derg *s.n.*	MK748452
L95	*T. baccata*	Clorhane *s.n.*	MK748453
L106	*T. baccata* 'Fastigiata'	Florence Court *s.n.*	_
L108	*T. baccata 'Amersfort'*	Mount Usher *s.n.*	_
L112	*T. baccata* 'Fructo Luteo'	Mount Usher *s.n.*	_
L145	*T. canadensis*	JFK Arboretum, New Ross 8.M.5	_
L149	*T. brevifolia*	JFK Arboretum, New Ross 33	MK748454
L162	*T. floridana*	Smith College Arboretum PULOG 7	MK748456
L164	*Austrotaxus spicata*	Edinburgh CAGNC*69	MK748455
L165	*T. canadensis*	Wendall Massachusetts *s.n.*	_
T1	*T. baccata*	Physic garden, TCD, *s.n.*	_
TH1	*T. canadensis*	Quebec, Canada	**MK685277**
TH2	*T. canadensis*	Quebec, Canada	**MK685278**

Glasnevin = National Botanic Gardens, Glasnevin, Dublin 9, Ireland

Edinburgh = Royal Botanic Gardens, Edinburgh, Scotland

JFK = John F. Kennedy Arboretum, New Ross, Co. Wexford, Ireland

Bedgebury = Bedgebury National Pinetum and Forest, Bedgebury, Kent, England

- Not submitted to GenBank

2.2. DNA extraction and purification

Total genomic DNA was extracted from 0.05-0.075g ground dry material or 0.075g-0.1g of fresh material using a modified hot CTAB method (Doyle and Doyle, 1987, Gawel and Jarret, 1991, Hodkinson et al., 2007). The crude DNA extract was then purified using JETquick Spin Columns (GENOMED Gmbh, Lohne, Germany) following the protocol outlined by GENOMED. Total genomic DNA was run on a 1.2% agarose gel infused with GelRed (Biotium, Hayward, California) at a volume of 3µl /80ml gel and viewed after electrophoresis on a Biometra Horizon gel rig under UV light on a DNR Bio-imaging or a BDA Biometra UVstar transilluminator. 6µl of total DNA was run for approximately 45mins at 125 volts.

2.3. PCR amplification

The column cleaned total genomic DNA was amplified using the polymerase chain reaction (PCR), (Mullis and Faloona, 1987) with a modification of the PCR method used by Hao et al. (2009). The PCR was optimised by adjusting the annealing temperature and the number of cycles, see section 2.7. KAPA Taq Ready Mix DNA Polymerase (Kapa Biosystems, Capetown South, Africa) and BIOLINE Biomix (Bioline reagents, United Kingdom) were used in the PCR reactions. PCR products were run on a 1.2% agarose gel to check for amplification. 3µl of PCR product were run for 20mins at 125 volts. All forward (F) and reverse (R) primers used in PCRs were obtained from the literature (Table 2.2) except for TS1(F), TS3(R), TS3(F) and TS5(R). These primers were manually designed by Dr. J Carolan (Maynooth University) from a Taxadiene synthase (*TS)* sequence alignment file for a number of Taxus species and accessions using BioEdit (Hall, 1999) to yield PCR products of approximately 600 to 700 base pairs when used together or in combination with published *TS* primers.

Table 2.2 Primers used for amplification of nrITS, trnL-F, TS and DBAT gene

Gene	Primer Code	Primer Sequence	Reference
ITS	AB101(F)	5'-ACGAATTCATGGTCCGGTGAAGTGTTC-3'	(Sun et al., 1994, Baldwin et al., 1995)
ITS	AB102(R)	5'-TAGAATTCCCCGGTTCGCTCGCCGTTA-3'	(Sun et al., 1994, Baldwin et al., 1995)
ITS	5.8	5'-GATGATTCACGGGATTCTG-3'	(Liston et al., 1996)
trnL	trnLc(F)	5'-CGAAATCGGTAGACGCTACG-3'	(Taberlet et al., 1991)
trnL	trnLf(R)	5'-ATTTGAACTGGTGACACGAG-3'	(Taberlet et al., 1991)
trnL	trnLd(F)	5'-GGGGATAGAGGGACTTGAAC-3'	(Taberlet et al., 1991)
trnL	Trnle(R)	5'-GGTTCAAGTCCCTCTATCCC-3'	(Taberlet et al., 1991)
TS	TS1-4(F)	5'-ATGGCTCAGCTCTCATTTAATGC-3'	(Hao et al., 2009)
TS	TS1-4(R)	5'-CGCAGCCGCCGAATTTGTCCA-3'	(Hao et al., 2009)
TS	TS5-9(F)	5'-TGGACAAATTCGGCGGCTGCG-3'	(Hao et al., 2009)
TS	TS5-9(R)	5'-CTTGTTGGAAGCTTCAACTCCTC-3'	(Hao et al., 2009)
TS	TS1(F)	5'-CATGGCGATCTGTGGCACC-3'	Designed by James Carolan*
TS	TS3(R)	5'-CACTCACTCTGTAAGCCTGG-3'	Designed by James Carolan*
TS	TS3(F)	5'-CACAGCCAAGTAGAACAAGG-3'	Designed by James Carolan*
TS	TS5(R)	5'-CAACCAGCGAAAGGCGTTCC-3'	Designed by James Carolan*
DBAT	DBAT(F)	5'-ATGGCAGGCTCAACAGAATTTG-'3	(Hao et al., 2009)
DBAT	DBAT(R)	5'-TCAAGGTTTAGTTACATATTTGTTTG-'3	(Hao et al., 2009)

*Designed by James Carolan but tested and developed further here.

The PCR product was purified using the Exosap method. The Exosap method cleans up any surplus dNTPs and primers which can interfere with the sequencing reaction. The exonuclease I removes any remaining single stranded primers or single stranded DNA, while the Shrimp Alkaline Phosphatase hydrolyses any unused dNTPs. Two cycle sequencing reactions were performed for each sample, one using the forward primer and the other using the reverse primer. The cycle sequencing reaction was run in a thermocycler (Veriti 96 well, Applied Biosystems) and then either stored at 4^{o}C or prepared to be run on the 3130Xl Genetic Sequencer (Applied Biosystems, Carlsbad, California).

To prepare the samples to be run on the Sequencer, the cycle sequencing products were cleaned using the Applied Biosystems Big Dye X-Terminator purification protocol. The cleaned samples were then loaded onto the sequencer and run on either a 36cm array

or an 80cm array with the manufacturer's protocol. The majority of samples were run on a 36cm array.

2.4. Cloning of PCR products

Some samples required cloning to give clearer sequences because the PCR product was heterogeneous. Cloning was performed using a Thermo Scientific CloneJET PCR cloning kit (Fermentas, Lithuania). The PCR product was inserted into the pJET1.2/blunt cloning vector which was then transformed into *E. coli* cells. The cells were then incubated and grown overnight at 37^0C. Eight single colonies where chosen randomly from the agar plate and a PCR was performed on each colony using the same primers as the initial pre-cloning amplification. A small part of the colony was picked directly from the agar and placed in the reaction using a sterile pipette tip. The resulting PCR product was run on an agarose gel to check which colonies cloned successfully. These samples were than sequenced following the same method as the standard sequencing reactions described above.

2.5. Leaf impressions

To help with the taxonomic identification of samples, leaf impressions were taken from a selection of samples to visualise the rows of stomata. This was a preliminary study. Clear nail varnish and Sellotape was used to create an impression of the abaxial side (underside) of the leaf on a slide. The leaf impressions were examined and photographed under a stereomicroscope at 10X and 20X magnification.

2.6. Data analysis

Sequences were edited and assembled in Geneious Pro 5.6.4 (Biomatters Ltd.). The sequences were then aligned in Geneious using either Muscle or Geneious algorithms with default settings. Some manual alignment was also required. Phylogenetic analyses were undertaken in MEGA 7 (Kumar et al., 2016) and MrBayes (Ronquist and Huelsenback, 2003). Haplotype networks were constructed using PopArt (Leigh and Bryant, 2015).

2.7. Protocols

Detailed protocols for each of these steps (DNA extractions, PCR amplification, Cycle sequencing, Cloning of PCR products, leaf impressions and data analysis) are given below:

2.7.1. Isolation and purification of total genomic DNA using CTAB

Materials:

2xCTAB Buffer (100mM Tris-HCl pH8.0, 1.4M NaCl, 20mM EDTA, 2% CTAB w/v; Sigma, Steinheim Germany).

2-mercaptoethanol (Sigma).

C:I (24:1 chloroform: isoamyl alcohol) (TCD supplies, Sigma).

Isopropanol (TCD supplies).

Wash buffer (70% ethanol) (TCD supplies).

TE Buffer (10mM Tris HCl pH8.0, 1mM EDTA) (Sigma).

PVP (Sigma)

Leaf material.

EDTA – Ethylenediaminatetraacetate acid disodium salt dihydrate (VWR, Belgium).

CTAB – Hexadecyltrimethylammonium bromide (Sigma).

PVP – Polyvinyl-pyrrolidone (Sigma).

Method:

1. Heat water bath (Grant, Cambridge) to 65^0C.

2. In a fumehood add 5ml of 2xCTAB, 20µl of mercaptoethanol and 0.1g of PVP to a 101x16.5mm capped centrifuge tube (Sarstedt, Ireland).

3. Place the tube in the water bath along with the mortar and pestle.

4. Weigh between 0.05-0.1g of leaf material.

5. Remove the mortar and pestle from water bath and dry with some tissue. Cut the sample leaf material into the pestle.

6. Grind the leaf using a small amount of the extraction buffer. When the leaf material is almost ground, add the remainder of the extraction buffer (and grind to an even slurry).

7. Pure the slurry back into a labelled capped centrifuged tube and incubate in the waterbath for 10mins mixing occasionally.

8. In the fume hood add 5ml of C:I.

9. Place on a horizontal shaker for 30mins. Afterwards open and close the lid to release pressure.

10. Spin in centrifuge (Thermo Electron Corporation IEC CL31R multispeed centrifuge) at 4,000rpm (approx. 3,500g) for 10mins.

11. Using a transfer pipette take off the top layer being careful not to take up any plant debris. Transfer into a 50ml conical based tube (Sarstedt).

12. Add an equal volume of isopropanol, about 5mls. Invert the tube gently to precipitate the DNA.

13. Place the samples in -20^{0}C freezer for approximately 1 week to further precipitate the DNA (overnight minimum).

14. Centrifuge the samples at 2,000rpm (approx. 1,500 g) for 5mins.

15. Pour off the supernatant in one smooth motion. The DNA pellet will stay at the bottom of the tube. Add 3ml of wash buffer to the tube and mix gently.

16. Centrifuge again at 2,000rpm for 5mins.

17. Pour off the ethanol and invert the tube on a piece of tissue for 5mins to drain away excess wash buffer.

18. Make sure all traces of ethanol are removed by turning the tube the right way up and allow to dry for approximately 30mins.

19. Re suspend pellet in 0.5ml 1xTE buffer and transfer to a labelled 1.5ml micro-centrifuge tube.

20. Store extracted DNA at -80^{0}C until required.

2.7.2. Purification of total DNA and cloned PCR products using spin columns

Materials:

JETquick Spin Columns (GENOMED Gmbh, Lohne, Germany).

Method: The protocol outlined by GENOMED was followed.

2.7.3. Preparation of 1.2 % agarose gel

Materials:

Agarose powder (Fisher Scientific, Fairlawn New Jersey).

TBE 10x (0.89M Tris, 0.89M Boric acid, 0.1M EDTA pH8.3) (Sigma).

1 x TBE Buffer (diluted from 10x stock).

Gel Red dye (Biotium, Hayward).

Method:

1. Weigh 1.2g of agarose powder.
2. Add to 100mls of 1xTBE buffer.
3. Heat in microwave for approximately 3-5mins to dissolve.
4. Pour 80mls into a 100ml Duran bottle.
5. Add 2µl of Gel Red dye.
6. When gel is cool enough to hold by hand pour into plastic well and insert combs.
7. Leave to set for approximately 30mins.

2.7.4. Gel electrophoresis to visualise total genomic DNA or PCR products

Materials:

TBE 10x (0.89M Tris, 0.89M Boric acid, 0.1M EDTA pH8.3).

1.2% agarose gel.

Loading dye (sucrose 40% w/v and bromophenol blue 0.25% w/v).

Method:

Prepare a 1.2% agarose gel and place in gel rig containing 1X TBE buffer.

Pipette 2µl of load dye onto a piece of Parafilm.

Pipette 2-6µl of total DNA or 3µl of PCR product directly onto the blob of loading dye and mix.

Pipette the mix into a well on the agarose gel.

Run the gel at 124volts for 30-45mins.

Place the agarose gel under UV transilluminator (Biometra, BDA compact UV Star) and photograph (Canon, EOS 1100D).

2.7.5. PCR amplification using Bioline BioMix or KAPA Taq Ready Mix

Materials:

Ultrapure water	5.375µl
Bioline Biomix or Kapa ready mix	6.25µl
Fprimer (20pmolµl^{-1})	0.125µl
Rprimer (20pmolµl^{-1})	0.125µl
DNA template (ca. 100ngµl^{-1})	0.625µl

Method:

1. Turn on thermocycler and set to appropriate program.
2. Thaw DNA and reagents on ice.
3. Made a master mix containing ultra pure water, biomix, forward and reverse primers.
4. Briefly vortex master mix.
5. Pipette 0.625µl template DNA into PCR tube and add 11.875µl of master mix.
6. Mix tube and spin down in centrifuge.
7. Place in thermocycler and run PCR program. Refer to conditions in table 2.4 -2.10.
8. Purify the PCR product using the Exosap method.
9. Check PCR amplified successfully by running PCR products on an electrophoresis gel.

2.7.6. PCR product purification: ExoSAP method for total 10µl reaction

Materials:

PCR product to be cleaned.

0.3µl of Exonuclease 20,000 U/ml Exonuclease I (Biolabs, New England).

2µl of Shrimp alkaline phosphatase (1U/µl) (Roche Diagnostics GmbH, Mannheim Germany).

7µl of Sterile ultra pure water.

Note: U=units of enzyme

Method:

1. Add 5µL of PCR product to tube.
2. Make ExoSAP master mix and mix thoroughly:
3. Add 5uL of ExoSAP master mix to tube containing PCR product.
4. Run on a thermal cycler set for the following program.

    37° for 30mins.

    82° for 20mins.

    4° for ∞

5. Store in freezer (-20^{o}C).

2.7.7. Cycle sequencing

A cycle sequencing reaction was performed on the purified PCR products using the following BigDye Terminator reaction mixture.

Materials:

5.5 µl sterile ultrapure water.

1µl of BigDye Terminator mix version 3.1 (Applied Biosystems).

1.5µl of 5x buffer (Applied Biosystems).

2µl of purified DNA (approximately 200ngµl^{-1}).

0.5µl of forward or reverse primer (20pmolµl^{-1}).

Method:

1. Add 2µl of purified DNA into a well of an applied biosystems plate and keep on ice.
2. Make up master mix for forward or reverse reaction.
3. Vortex master mix.
4. Add 8µl of master mix to each well containing DNA.
5. Seal plate.
6. Mix plate and spin briefly in centrifuge.
7. Place in thermocycler set to the appropriate program, Thermocycler BigDye kit standard program for cycle sequencing.

2.7.8. Cycle Sequencing Purification

Materials:

BigDye Xterminator Purification Kit (Applied Biosystems, Foster City California).

Method:

1. Add 45µl of SAM solution to each well.
2. Vortex the BigDye X terminator Solution extremely well before use and after every row of samples.
3. Add 10µl of BigDye X terminator Solution to each well, using a wide bore pipette tip or a regular tip with the end cut short.
4. Seal the plate with an adhesive cover.
5. Mix the plate for 30mins using a plate mixer set to 2,000rpm.
6. Centrifuge the plate at 1,000xg for 2mins.
7. The plate is now ready to run on the sequencer.

2.7.9. Cloning of PCR products

Materials:

Thermo Scientific CloneJET PCR Cloning Kit (Fermentas, Lituania).

Room temperature LB broth pH7.0 (10g Tryptone, 5g Yeast extract, 5g NaCl ; Sigma). LB broth adjusted to pH with NaOH.

Agar plates (LB broth, 15g agar; Sigma, 2ml ampicillin 50 mg/ml; GiBCO).

PCR product to be cloned.

Method:

Follow the sticky-end cloning protocol as in the Thermo Scientific CloneJET PCR Cloning Kit

Stage I: Blunting reaction

Set up on ice the following:

2x reaction buffer 10µl

PCR product: 1µl if strong or 2µl if weak

Ultra pure water 5µl-6µl depending on how much PCR product used

DNA blunting enzyme 1µl

1. Add all together in a small tube to give a total volume of 18µl.

2. Vortex for 3-5secs.

3. Incubate at 70^0C for 5mins and then chill on ice.

Ligation reaction

Set up on ice the following:

pJET 1.2 1µl

T4 DNA ligase 1µl

5. Add pJET and T4 DNA ligase to blunting reaction for each sample to give a total volume of 20µl.

6. Vortex and centrifuge for 3-5secs.

7. Incubate at room temperature 22^0C for 5mins.

8. Use ligation mix directly for transfer to competent cells or store in a freezer at - 20^0C.

Stage II: Transformation of competent cells

1. Thaw competent cells on ice.
2. Add 2µl from Stage I to competent cells and mix gently by tapping. Do NOT vortex.
3. Leave on ice for 5mins (can be left up to a maximum of 30mins).
4. Heat shock the cells for 30-60secs at 42^0C. Do NOT shake.
5. Place on ice for 1min.
6. Add 250µl of LB broth.
7. Pipette 50µl of this mixture onto an agar plate and spread by rolling around 5-15 glass beads.
8. Clean glass beads for reuse in 100% ethanol.
9. Incubate agar plates at 37^0C for 16 hours.
10. Choose at random eight single colonies from each agar plate. Pick out the single colony using a pipette tip and transfer into a PCR tube containing 10µl of LB broth.
11. Perform PCR as per cloned PCR protocol.

2.7.10. Cycle sequencing of cloned PCR products

Materials:

Ultrapure water	5.5µl
Pink Mix (Big Dye Terminator v.3.1)	1.0µl
5x Buffer	1.5µl
Forward or Reverse Primer (4pmolμl^{-1})	0.5µl
PCR product	1.5µl

Method:

1. Turn on the thermocycler and choose “BigDye Kit standard program”.
2. Thaw reagents on ice.
3. Make up a master mix containing ultrapure water, pink mix, 5x Buffer and forward primer.
4. Shake the master mix and pulse in centrifuge.
5. Add 8.5µl of the master mix to 1.5µl of cloned PCR product.
6. Repeat the process for the reverse primer.

7. Seal plate and insert into thermocycler.

8. Run program.

Thermocycler Big dye kit standard program for cycle sequencing

Initial denaturation	96^0 C x 1min
Followed by 25 cycles of:	96^0 C x 10secs
50^0 C x 5secs	
60^0 C x 4mins	
Hold	4^0 C ∞

2.7.11. Cycle sequencing for trnL-F

A cycle sequencing reaction was performed on the purified PCR products using the following BigDye Terminator reaction mix.

Materials:

Ultrapure water.

BigDye Terminator mix version 3.1 (Applied Biosystems).

5x buffer (Applied Biosystems).

1.5µl of purified DNA (approximately 200 $ng\mu l^{-1}$).

Forward or reverse primer ($10pmol\mu l^{-1}$).

Master mix:

Sterile ultrapure water	5.8µl
BigDye Terminator mix	0.8µl
5x buffer	1.7µl
F or R primer ($10pmolul^{-1}$)	0.2µl

Method:

1. Take 1µl of the Exosap product and add to 9µl of purified water.

2. Add 1.5µl of the diluted exosap product DNA into a well of an applied biosystems plate and keep on ice.

3. Make up master mix for forward or reverse reaction.

4. Vortex master mix.

5. Add 8.5µl of master mix to each well containing DNA.

6. Seal plate.

7. Mix plate and spin briefly in centrifuge.

Place in Thermocycler set to the appropriate program, Thermocycler Bigdye kit standard program for cycle sequencing.

2.7.12. Cycle sequencing for ITS 5.8

A cycle sequencing reaction was performed on the purified PCR products using the following BigDye Terminator reaction mix.

Materials:

3.5µl ultrapure water.

0.8µl of BigDye Terminator mix version 3.1 (Applied Biosystems).

1.7µl of 5x buffer (Applied Biosystems).

1.5µl of purified DNA (approximately 200ngµl^{-1}).

2.5µl Forward or reverse primer (10ngµl^{-1}).

Method:

1. Add 1.5µl of purified DNA (approximately 200ngµl^{-1}) into a well of an Applied Biosystems plate and keep on ice.

2. Make up master mix for forward or reverse reaction.

3. Vortex master mix.

4. Add 8.5µl of master mix to each well containing DNA.

5. Seal plate.

6. Mix plate and spin briefly in centrifuge.

7. Place in thermocycler set to the appropriate program, thermocycler BigDye kit standard program for cycle sequencing.

2.7.13. Cycle sequencing for ITS AB101 and AB102

A cycle sequencing reaction was performed on the purified PCR products using the following BigDye Terminator reaction mix.

Materials:

5.5µl ultrapure water.

0.8µl of BigDye Terminator mix version 3.1 (Applied Biosystems).

1.7µl of 5x buffer (Applied Biosystems).

2.0µl of purified DNA (approximately 200 $ng\mu l^{-1}$).

0.5µl forward or reverse primer (20$pmol\mu l^{-1}$).

Method:

1. Add 2µl of purified DNA into a well of an applied biosystems plate and keep on ice.
2. Make up master mix for forward or reverse reaction.
3. Vortex master mix.
4. Add 8µl of master mix to each well containing DNA.
5. Seal plate.
6. Mix plate and spin briefly in centrifuge.

Place in thermocycler set to the appropriate program, thermocycler BigDye kit standard program for cycle sequencing.

Table 2.3. Cycle Sequencing Parameters

		Water (µl)	Big Dye (µl)	5xBuffer (µl)	Forward primer or Reverse primer	Master Mix (µl)	Exosap product (diluted 10 times)
trnL- F		5.8	0.8	1.7	0.2 µl (10pmol/µl)	8.5	1.5
5.8 *ITS*		3.5	0.8	1.7	2.5ul (10ng/µl)	8.5	1.5
5.8 *ITS*		3	1	1.5	3ul (10ng/µl)	8.5	1.5
AB101&AB102 *ITS*		5.5	1	1.5	0.5ul (10pmol/µl)	8.5	1.5
AB101&AB102 *ITS*		5.5	0.8	1.7	0.5ul (10pmol/µl)	8.5	1.5
TS		5.5	1	1.5	0.5ul (20pmol/µl)	8.5	1.5
DBAT		5.5	1	1.5	0.5ul (20pmol/µl)	8.5	1.5
DBAT		5.5	1	1.5	0.5ul (4pmol/µl)	8	2
TS outgroups		5.5	1	1.5	0.5ul (4pmol/µl)	8.5	1.5

Table 2.4. PCR Parameters

	Water (µl)	Biomix (µl)	Forward and Reverse primer	Magnesium (µl)	Mater Mix (µl)	DNA (µl)
trnL-F	4.625	6.25	0.5µl (10pmol/µl)		12.2	0.3
TS 1F3R &3F5R	5.375	6.25	0.125µl (20pmol/µl)		11.875	0.625
*TS*14	5.375	6.25	0.125µl (20pmol/µl)		11.875	0.625
*TS*59	5.375	6.25	0.125µl (20pmol/µl)		11.875	0.625
TS out - groups	5.25	6.25	0.125µl (20pmol/µl)	0.125ul	11.875	0.625
ITS clones	4.75	6.25	0.5µl (10pmol/µl)		12	0.5
ITS	4.95	6.25	0.5µl (10pmol/µl)		12.2	0.3
DBAT	5.375	6.25	0.125µl (20pmol/µl)		11.875	0.625

Table 2.5. Thermocycler parameters for trnL-F – cycles 30

	^{o}C	Time
Premelt	95	1min 30sec
Denaturation	95	45sec
Annealing	50	45sec
Extending	72	2min
Final extending	72	7min
Pause	4	∞

Note: The following samples had an annealing temperature of 52^{o}C P29, P30, L93, L94, L95, L149, L162, L164

Table 2.6 Thermocycler parameters for TS – cycles 38

	^{o}C	Time
Premelt	95	1min
Denaturation	95	30sec
Annealing	59	1min
Extending	72	1min 30sec
Final extending	72	1min
Pause	4	∞

Note: Annealing temperature for different genera; *Podocarpus* 55 ^{o}C and *Torreya* 50 ^{o}C

Table 2.7. Thermocycler parameters for cloned TS – cycles 29

	°C	Time
Premelt	94	10min
Denaturation	95	30sec
Annealing	59	1min
Extending	72	1min 30sec
Final extending	72	10min
Pause	4	∞

Table 2.8. Thermocycler parameters for DBAT – cycles 38

	°C	Time
Premelt	95	1min
Denaturation	95	45sec
Annealing	49	4min
Extending	72	2min
Final extending	72	7min
Pause	4	∞

Table 2.9. Thermocycler parameters for TS1F3R and 3F5R– cycles 38

	°C	Time
Premelt	95	1min
Denaturation	95	30sec
Annealing	55	1min
Extending	72	1min 30sec
Final extending	72	7min
Pause	4	∞

Table 2.10 Thermocycler parameters for ITS – cycles 32

	°C	Time
Premelt	94	90sec
Denaturation	94	30sec
Annealing	62	1min
Extending	72	1min
Final extending	72	7min
Pause	4	∞

2.7.14. Leaf impressions

Materials:

Clear nail varnish.

Sellotape.

Microscope slide.

Microscope with camera Leica ICC50HD.

Method:

1. Apply a thin layer of clear nail varnish to the underside of the leaf.
2. Allow to dry.
3. Apply a piece of Sellotape over the dried nail varnish.
4. Peel the Sellotape off the leaf.
5. Smooth the Sellotape onto a clean microscope slide.
6. Take photos of the leaf impressions using a camera attached to microscope.

Chapter 3: Phylogenetics of *Taxus* using the internal transcribed spacers of nuclear ribosomal DNA and plastid *trnL-F* regions

Phylogenetics of *Taxus* using the internal transcribed spacers of nuclear ribosomal DNA and plastid *trnL-F* regions

Published in Journal Horticulturae (doi.org/10.3390/horticulturae6010019)

Coughlan P[1], Carolan JC[2], Hook IL[3], Kilmartin L[1] & Hodkinson TR[1]

[1]Botany Building, School of Natural Sciences, Trinity College Dublin, the University of Dublin, Ireland

[2] Department of Biology, Maynooth University, Maynooth, Co. Kildare

[3]School of Pharmacy and Pharmaceutical Sciences, Trinity College Dublin, the University of Dublin, Ireland

3.1. Abstract

Taxus L. is a genus of trees and shrubs with high value in horticulture and medicine as a source of the anticancer drug paclitaxel. The taxonomy of the group is highly complex due to the lack of diagnostic morphological characters and the high degree of similarity among species. *Taxus* has a wide global geographic distribution and some taxonomists recognise only a single species with geographically defined subgroups whereas others have described several species. To address these differences in taxonomic circumscription a phylogenetic analysis was conducted involving representatives of the different species and subspecies currently recognised. Maximum Liklihood, Bayesian Inference and TCS haplotype network analyses were conducted on DNA sequences obtained for the nuclear internal transcribed spacer (*ITS*) and the plastid gene regions *trnL* intron and *trnL-F* intergenic spacer (*trnL-F*). The results support the recognition of nine distinct species: *T. baccata, T. brevifolia*, *T. canadensis*, *T. cuspidata*, *T. floridana*, *T. fuana*, *T. globosa, T. sumatrana* and *T. wallichiana* but evidence is found for less species distinction and considerable reticulation within the *T. baccata*, *T. canadensis* and *T. cuspidata* group. Evidence is also presented for the sister group status of *Pseudotaxus* to *Taxus* and the inclusion of *Amentotaxus*, *Austrotaxus*, *Cephalotaxus* and *Torreya* within the Taxaceae. We compare the results to known taxonomy, present new leaf anatomical data and discuss the origins of the hybrids *T.* x *media* and *T.* x *hunnewelliana.*

3.2. Introduction

Morphological differences between species of *Taxus* L. are slight and individuals within species have high levels of phenotypic plasticity, their morphology varying considerably with the environment (Collins et al., 2003). This often leads to misidentification and problems with classification using traditional taxonomic methods that rely solely on morphological characters. Indeed, delimitation of species in *Taxus* has been a long-standing taxonomic problem. Pilger (1903) and Elwes and Henry (1906) classified *Taxus* as monotypic with several geographical subspecies, a view supported by Dempsey and Hook (2000) who provided evidence based on needle morphological variation and chemical characteristics. Furthermore, Dempsey (2000) conducted a morphological and phytochemical analysis to investigate intra- and inter-generic relationships of *Taxus* and related genera and concluded that few characters could be used to distinguish among *Taxus* species but that a clear distinction between the sister taxa *Cephalotaxus* and *Torreya* could be resolved.

Contrary to this monotypic view of *Taxus,* numerous authors have recognised multiple species including Pilger (1903), Cope (1998), Farjon (1998) who resolved 7-12 species, while Spjut (2007a) recognised 24 species and 55 varieties. Currently nine species are listed in the Plant List (2013) namely *T. baccata* L. (Europe, Canada), *T. brevifolia* Nutt. (western North America), *T. canadensis* Marshall (eastern Canada, USA), *T. cuspidata* Siebold & Zucc. (Japan), *T. floridana* Nutt. ex Chapm. (Florida, south United States), *T. fuana* Nan Li & R.R. Mill, *T. globosa* Schltdl (Mexican), *T. sumatrana* (Miq.) de Laub. (China) and *T. wallichiana* Zucc. (Eastern India) (Figure 3.1).

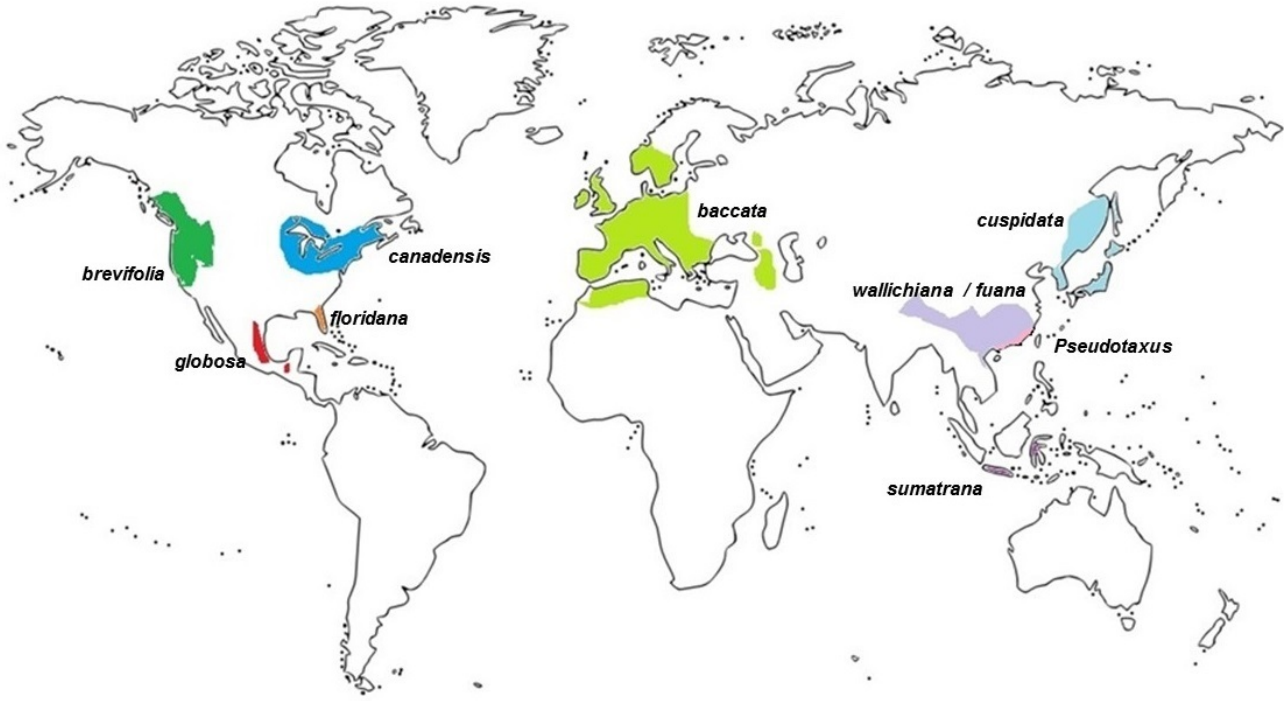


Figure 3.1. Distribution of Taxus species and its sister genus Pseudotaxus adapted from Li et al. (2001)

The importance of species delimitation and classification of *Taxus* goes beyond taxonomy and comparative biology. The genus comprises members that synthesise the anti-cancer taxane drug paclitaxel (Taxol) and incorrect species identification can hinder cultivation and drug production efforts (Appendino, 1995, Dempsey and Hook, 2000). Phytochemical studies examining plants collected from different regions of the world, require a stable nomenclature and species identification. Furthermore, an accurate phylogenetic reconstruction is required to infer how paclitaxel may have evolved and is synthesised which may have biotechnological implications.

Accurate species identification is also required for conservation of *Taxus* species and forests (Lanker et al., 2010). According to the International Union for Conservation of Nature (ICUN) Red List, several species of *Taxus* are threatened to different degrees (Thomas and Farjon, 2011, Spector et al., 2011, Thomas et al., 2013, Thomas, 2013b, Thomas, 2013a, Yang et al., 2013). *Taxus brevifolia* and *Taxus mairei* are classified as Near Threatened and Vulnerable, respectively, with their numbers currently decreasing. A major part of this decline is attributed to logging (Thomas, 2013a) and the harvesting of bark for paclitaxel production although this exploitation has largely stopped following the development of a semi synthetic process to produce paclitaxel from *T. baccata.*

Taxus wallichiana, T. globosa, *T. chinensis* and *T. fuana* (*T. contorta*) are all classified as Endangered (Thomas and Farjon, 2011, Thomas, 2013b, Thomas et al., 2013, Thomas, 2011). This is due to the overexploitation of *T. walliachiana* and *T. chinensis* for paclitaxel production. Deforestation has caused populations of *Taxus globosa* to decline, *T. globosa* is not yet exploited for paclitaxel production. *Taxus fuana* according to the Red List has become endangered due to over exploitation associated with medical use along with overcollection for fuel and fodder. Shah et al. (2008a) observed that the populations of *T. fuana* were declining mainly due to human pressure from habitat destruction, deforestation and overexploitation for fuel, fodder, timber and farming. The most endangered species is *Taxus floridana* which is listed as Critically Endangered (Spector et al., 2011).

Molecular phylogenetic approaches offer the potential to better understand and resolve the relationships within taxonomically difficult groups such as *Taxus* and a number of studies exist that provide insight into the evolution and phylogenetics of the genus and its relatives. Chaw et al. (1993) and Cheng et al. (2000) used DNA sequences of 18S ribosomal DNA, internal transcribed spacer regions of nuclear ribosomal DNA and *matK* plastid DNA to clarify the phylogenetic position of Taxaceae in the gymnosperm order Taxales. There is also evidence for the monophyly of *Taxus* (Chaw et al., 1993, Hao et al., 2008a) but the relationships of its species remain understudied and not fully resolved (Chaw et al., 1993, Cheng et al., 2000, Li et al., 2001, Hao et al., 2008a). Extant yews are believed to have evolved from a group including the fossil *Paleotaxus redivia* (Appendino, 1993). Triassic *P. redivia* existed over 200 million years ago and is believed to be the oldest yew according to fossil records (Appendino, 1993). A mid–Jurassic relative (140 myr old) is said to be more recognizable as a member of *Taxus* and hence named *T. jurassica* (Florin, 1951). A Quaternary fossil yew, *T. grandis,* is probably a synonym of *T. baccata* (Hartzell, 1991).

The aim of this study was to examine the phylogenetic relationships within *Taxus* by comparing nucleotide sequences obtained from the internally transcribed spacer region (*ITS*) of 18S-26S nuclear ribosomal DNA and the *trnL* intron and the *trnF* intergenic spacer region of plastid DNA (*trnL-F*). These regions have been used extensively in

systematic studies of other groups of plants for investigating relationships at different taxonomic levels (Taberlet et al., 1991, Sun et al., 1994, Baldwin et al., 1995, Wendel et al., 1995, Li et al 2001, Hodkinson et al., 2002, Carolan et al., 2006, Hao et al., 2008b) and have also been applied to limited samples of *Taxus*. Representatives of *Austrotaxus*, *Amentotaxus*, *Cephalotaxus*, *Pseudotaxus*, and *Torreya* were also included to assess inter-relationships in Taxaceae and related families (Cephalotaxaceae and Amentotaxaceae). A large-scale phylogenetic analysis of the group including all currently accepted *Taxus* species according to the Plant List (2013) was conducted and we also examine evidence for evolutionary reticulation using network analyses and a detailed assessment of the putative hybrid taxa, *T.* x *media* and *T.* x *hunnewelliana*.

3.3. Materials and Methods

3.3.1. Specimens

Fresh plant material was obtained from the National Botanic Garden, Glasnevin, Ireland. DNA samples were also available from the Trinity College Dublin (TCD) DNA Bank. Voucher specimens were kept for each sample, dried and stored in the TCD Herbarium. Some ground and dried leaf samples were also available from our previous work on the genus (Table 3.1).

Table 3.1. Samples newly sequenced for ITS and trnL intron and trnL-F intergenic spacer or used in the morphological assessments

	Name	Voucher or living specimen, Accession ID	GenBank Code *ITS*; *trnL-F*
P1	*T. canadensis*	Glasnevin Hook *s.n.*	**MK116527**; MK731927
P2	*T. canadensis* 'Aurea'	Glasnevin Hook *s.n.*	**MK123469**; MK731928
P3	*T. cuspidata*	JFK Hook *s.n.*	**Mk116528**
P5	*T. canadensis*	JFK Hook *s.n.*	**MK116529**; MK731929
P6	*T. canadensis*	Canada Hook *s.n.*	**MK116530**; MK731930
P8	*Cephalotaxus harringtonia*' Fastigiata'	XX.006542 Glasnevin Coughlan P8	**MK116531**; MK731931
P9	*T. cuspidata f. thayerae*	1952.006540 Glasnevin Coughlan P9	**MK168608, MK168609**; MK731932
P10	*T. canadensis* 'Aurea'	XX.006549 Glasnevin Coughlan P10	**MK168610, MK168611, MK168612**; MK731933
P11	*T. canadensis*	XX.006556 Glasnevin Coughlan P11	**MK168613, MK168614, MK168615**; MK731934
P12	*T. cuspidata var. luteobaccata*	1930.006571Glasnevin Coughlan P12	**MK168616, MK168617, MK168618**; MK731935
P13	*T.* 'Aurea'	XX.006570 Glasnevin Coughlan P13	MK731936
P14	*T. brevifolia*	1885.006579 Glasnevin Coughlan P14	**MK123470**; MK731937
P15	*T. cuspidata*	XX.006597 Glasnevin Coughlan P15	**MK123471**; MK731938
P16	*T. x media* 'Hicksii'	XX.006589 Glasnevin Coughlan P16	**MK123472**; MK731939
P17	*T. cuspidata*	1911.006591 Glasnevin Coughlan P17	**MK123473**; MK731940
P18	*T. cuspidata f. thayerae*	1952.006593 Glasnevin Coughlan P18	**MK168781, MK168782, MK168783**; MK731941
P19	*T. cuspidata*	XX.006601 Glasnevin Coughlan P19	**MK168784, MK168785, MK168786**; MK731942
P20	*T. canadensis* 'Aurea'	XX.006609 Glasnevin Coughlan P20	MK731943
P21	*T. canadensis*	XX.005502 Glasnevin Coughlan P21	**Mk21148, MK211149, MK211150**; MK731944
P22	*T. x media* 'Cuftoni'	XX.0055077 Glasnevin Coughlan P22	**MK211151**; MK731945
P23	*Podocarpus salignus*	XX.007572 Glasnevin Coughlan P23	MK731946
P24	*Torreya californica*	XX.007746 Glasnevin Coughlan P24	**MK211152**; MK731947
P25	*Unlabelled Torreya Sp.*	XX.007599 Glasnevin Coughlan P25	**MK211153**; MK731948
P26	*Podocarpus macrophyllus*	2005.0135 Glasnevin Coughlan P26	MK731953
P27	*T. baccata f.aurea*	XX.006574 Glasnevin Coughlan P27	**MK211154, MK211155, MK211156**; MK731949
P28	*T. baccata* 'Fastigiata'	13638 Glasnevin Coughlan P28	**MK211157**; MK731950
P29	*T. baccata* 'Grandis'	2001.4181 Glasnevin Coughlan P29	MK731951

P30	*T. baccata* 'Fastigiata'	Glasnevin Cemetery Coughlan P30	MK731952
L10	*T. baccata*	Ranelagh Park, Dublin	MK783697
L24	*T. canadensis*	Bedgebury 14/94	**MK272737**; MK748448
L25	*T. x hunnewelliana*	Bedgebury 13/125	MK783698
L26	*T. x media* 'Hicksii'	Bedgebury 13/006	MK783699
L27	*T. cuspidata*	Bedgebury 13/136	MK783700
L31	*T. brevifolia*	Glasnevin 1885.006579	MK783701
L32	*T. cuspidata* 'Fructo luteo'	Glasnevin 12/30Kew	MK783702
L33	*T. canadensis*	Glasnevin XX.006556	**MK272738**; MK748449
L34	*Torreya nucifera*	Glasnevin XX.007734	**MK272739**
L36	*T. canadensis* 'Aurea'	Glasnevin XX.006549	**MK272740**; MK783703
L42	*T. baccata* 'Fastigiata'	Bedgebury 13/108	MK783704
L40	*T. x media* 'Hatfieldii'	Bedgebury 13/280	-
L44	*T. baccata* 'Rushmoor'	Bedgebury 15/0287	MK783705
L51	*T. baccata*	Bedgebury 13/025	MK783706
L66	*Amentotaxus formosana*	Edinburgh 19763745	MK748450
L68	*Cephalotaxus sinensis*	Edinburgh 19081025	-
L73	*Torreya jackii*	Edinburgh 19970112	-
L75	*Cephalotaxus oliveri*	Edinburgh 951689	-
L93	*T. baccata*	Cornalack Lough Derg *s.n.*	MK748451
L94	*T.baccata*	Cornalack lough Derg *s.n.*	MK748452
L95	*T. baccata*	Clorhane *s.n.*	MK748453
L106	*T. baccata* 'Fastigiata'	Florence Court *s.n.*	-
L108	*T. baccata 'Amersfort'*	Mount Usher *s.n.*	-
L112	*T. baccata* 'Fructo Luteo'	Mount Usher *s.n.*	-
L145	*T. canadensis*	JFK Arboretum, New Ross 8.M.5	-
L149	*T. brevifolia*	JFK Arboretum, New Ross 33	MK748454
L162	*T. floridana*	Smith College Arboretum PULOG 7	MK748456
L164	*Austrotaxus spicata*	Edinburgh CAGNC*69	MK748455
L165	*T. canadensis*	Wendall Massachusetts *s.n.*	-
T1	*T. baccata*	Physic garden, TCD, *s.n.*	-
TH1	*T. canadensis*	Quebec, Canada	**MK685277**
TH2	*T. canadensis*	Quebec, Canada	**MK685278**

Glasnevin = National Botanic Gardens, Glasnevin, Dublin 9, Ireland

Edinburgh = Royal Botanic Gardens, Edinburgh, Scotland

JFK = John F. Kennedy Arboretum, New Ross, Co. Wexford, Ireland

Bedgebury = Bedgebury National Pinetum and Forest, Bedgebury, Kent, England

- Not submitted to GenBank

3.3.2. Outgroup selection

Austrotaxus, Amentotaxus, *Cephalotaxus*, *Pseudotaxus* and *Torreya* were selected as outgroups based on previous studies. *Cephalotaxus* was sister to *Taxus* in Cheng et al. (2000) and Hao et al. (2008b) and Taxaceae comprises *Amentotaxus, Taxus* and *Torreya* (Farjon, 1998). *Cephalotaxus* is sometimes included in Taxaceae (Li and Fu, 1997).

3.3.3. DNA extraction and sequencing

DNA was extracted from 0.05-0.075g ground leaf material or 0.075-0.1g of fresh material using a modified hot CTAB method of Doyle and Doyle (Doyle and Doyle, 1987, Gawel and Jarret, 1991, Hodkinson et al., 2007). DNA was precipitated using 100% isopropanol, pelleted and washed with 70% ethanol and purified using the JETquick Spin Columns (GENOMED Gmbh, Lohne, Germany) following the manufacturer's instructions. DNA was then stored in TE buffer (10mM Tris/HCL, 1mM EDTA, pH8.0) at -80°C until required.

The forward and reverse primers of Sun et al. (1994) were used for amplification and sequencing of the *ITS* region. The internal primer 5.8 of Liston et al. (1996) was also used as a sequencing primer because of the long length of the amplicons. The *trn*L intron and the *trn*F spacer (hereafter the *trnL-F* region) were amplified and sequenced as one segment using primers "c" and "f" of Taberlet et al. (1991) and internal *trnL-F* primers "e" and "d" when necessary. PCRs for both regions were carried out in 12.5µl reactions using BIOLINE Biomix (Bioline reagents, UK). Both the *trnL–F* and *ITS* PCR amplifications were prepared using 4.95µl ultrapure water, 6.25µl Biomix, 0.5µl (10pmol) forward and reverse primers and 0.3µl of column-cleaned total DNA (ca. 100ngμl^{-1}). The reaction conditions for *trnL-F* were as follows: denaturation at 95°C for 1min 30sec followed by 30 cycles of 45sec at 95°, 45sec at 50°C, 2min at 72°C and a final extension at 72°C for 7min in an Applied Biosystems Verti 96 well thermal cycler. The reaction conditions for *ITS* were as follows: denaturation at 94°C for 1min 30sec followed by 32 cycles of 30sec at 94°, 1min at 62°C, 1min at 72°C and a final extension at 72°C for 7min in an Applied Biosystems Verti 96 well thermal cycler. 3µl PCR products were run on a 1.2% agarose gel to check for amplification. Successfully amplified DNA fragments were purified using the ExoSap method. 0.3µl of Exonuclease, 2µl of Shrimp Alkaline phosphatase (1U/µl)

and 7µl of sterile ultrapure water was added to 5µl of PCR product and incubated at 37° for 30mins followed by 82° for 20mins. Purified PCR products were then sequenced on an Applied Biosystems 3130xl genetic sequencer according to the manufacturers protocol with the same primers used for initial amplification.

3.3.4. DNA cloning

Some samples required cloning due to the heterogeneity of the PCR product. Cloning was performed using a Thermo Scientific CloneJET PCR cloning kit (Fermentas, Lithuania). The PCR product was inserted into the pJET1.2/blunt cloning vector that was then transformed into *E. coli* cells. The cells were incubated and grown overnight at 37°C. Eight single colonies where chosen randomly from the agar plate and a PCR was performed on each colony using the same primers as the initial pre-cloning amplification. A small part of the colony was picked directly from the agar and placed directly in the reaction using a sterile pipette tip. The parameters for the PCR were the same as above with the exception that there was an additional 10mins pre-melt at 94°C and 10mins final extension at 72°C. The cloned PCR products were purified using JETquick spin columns and sequenced as described above.

3.3.5. Leaf impressions

To help with the taxonomic identification of samples, leaf impressions were taken from a selection of samples to visualise the rows of stomata. The epidermal cell patterns were visualized using cellulose acetate impressions according to the method of Sarvella et al. (1961). Clear nail varnish and Sellotape were used to create an impression of the abaxial side of the leaf on a slide. The leaf impressions were examined and photographed under a stereomicroscope at 10X and 20X magnification. This was a preliminary study carried out on some samples of *T. cuspidata, T. canadensis, T. brevifolia, T. baccata and T. x media.*

3.3.6. Sequence analysis and phylogenetic reconstruction

To obtain a contiguous sequence for the target DNA region, forward and reverse sequence reads were assembled in Geneious Pro 5.6.4. (Biomatters Ltd.). The sequences were aligned in Geneious using highest sensitivity, using either Muscle or Geneious algorithms with default settings. The sequences were then manually aligned if necessary. The aligned matrix was imported into MEGA 7 (Kumar et al., 2016) for Maximum Likelihood (ML) phylogenetic analyses (Kumar et al., 2016) and also MrBayes (Ronquist and Huelsenback, 2003) for Bayesian phylogenetic tree reconstruction. Gaps smaller than 10bp were coded as missing data, unless they were found in regions where there was an obvious tandemly arranged duplication in one sequence that was clearly due to a single mutation (a duplication). Such duplications were scored as only one character in the subsequent phylogenetic analyses. Gaps larger than 10bp were excluded from the analyses.

For the ML analyses, the best fit substitution model was determined by the Model Selection function in MEGA7 (Kumar et al., 2016) and was found to be the Tamura 3-parameter model (T92) with gamma distributed rate heterogeneity and estimated proportion of invariant sites (G+I) for both the *ITS* and the *trnL-F* gene regions (Supplementary Tables 3.1, 3.2). ML was performed in MEGA7 with 1,000 replicates of random sequence addition and nearest neighbour interchange (NNI) branch swapping. Initial tree(s) for the heuristic search were obtained automatically by applying Neighbour-Joining and BioNJ algorithms to a matrix of pairwise distances estimated using the Maximum Composite Likelihood (MCL) approach, and then selecting the topology with superior log likelihood value. Bootstrap support was calculated from 1,000 replicates with the same settings as the initial search.

Bayesian inference of phylogeny was performed using MrBayes version 3 (Ronquist and Huelsenback, 2003). The T92+G +I best-fit nucleotide substitution model was used as determined for the ML analyses above. Four parallel Markov chain Monte Carlo (MCMC) chains were run for 25,000,000 generations with trees sampled every 1,000 generations, and 25% of trees were discarded as burn-in.

Haplotype networks were constructed using PopArt (Leigh and Bryant, 2015). Sequences were coded according to taxon name (trait) and haplotype networks constructed using TCS following Clement et al. (2002).

3.4. Results

The aligned *trnL–F* matrix, with unalignable regions removed, was 486bp long. Phylogenetic analyses using ML and Bayesian Inference (BI) were largely congruent. The ML bootstrap (BS) values are therefore shown on the Bayesian tree with the Bayesian posterior probability (PP) values (Figure 3.2; ML tree in Supplementary Figure 3.1). Analysis including all genera of Taxaceae clearly supports the monophyly of the Taxaceae genera (*Amentotaxu*s PP=1.0; *Austrotaxus* PP=1.0; *Cephalotaxus* PP=1.0; *Pseudotaxus* PP=1.0; *Taxus* PP=1.0 and *Torreya* PP=1.0). The sister status of *Pseudotaxus* to *Taxus* is strongly supported when the tree is mid-point rooted or rooted on any of the remaining genera. Despite high support for the monophyly of genera, the relationships among genera especially *Austrotaxus*, *Amentotaxus*, *Cephalotaxus* and *Torreya* are not well supported. Phylogenetic analysis of a reduced *trnL–F* dataset including only *Taxus* and *Pseudotaxus* was undertaken to examine infrageneric patterns in *Taxus* (Supplementary Figure 3.1b). Some species groupings are evident within *Taxus* but these are not strongly supported.

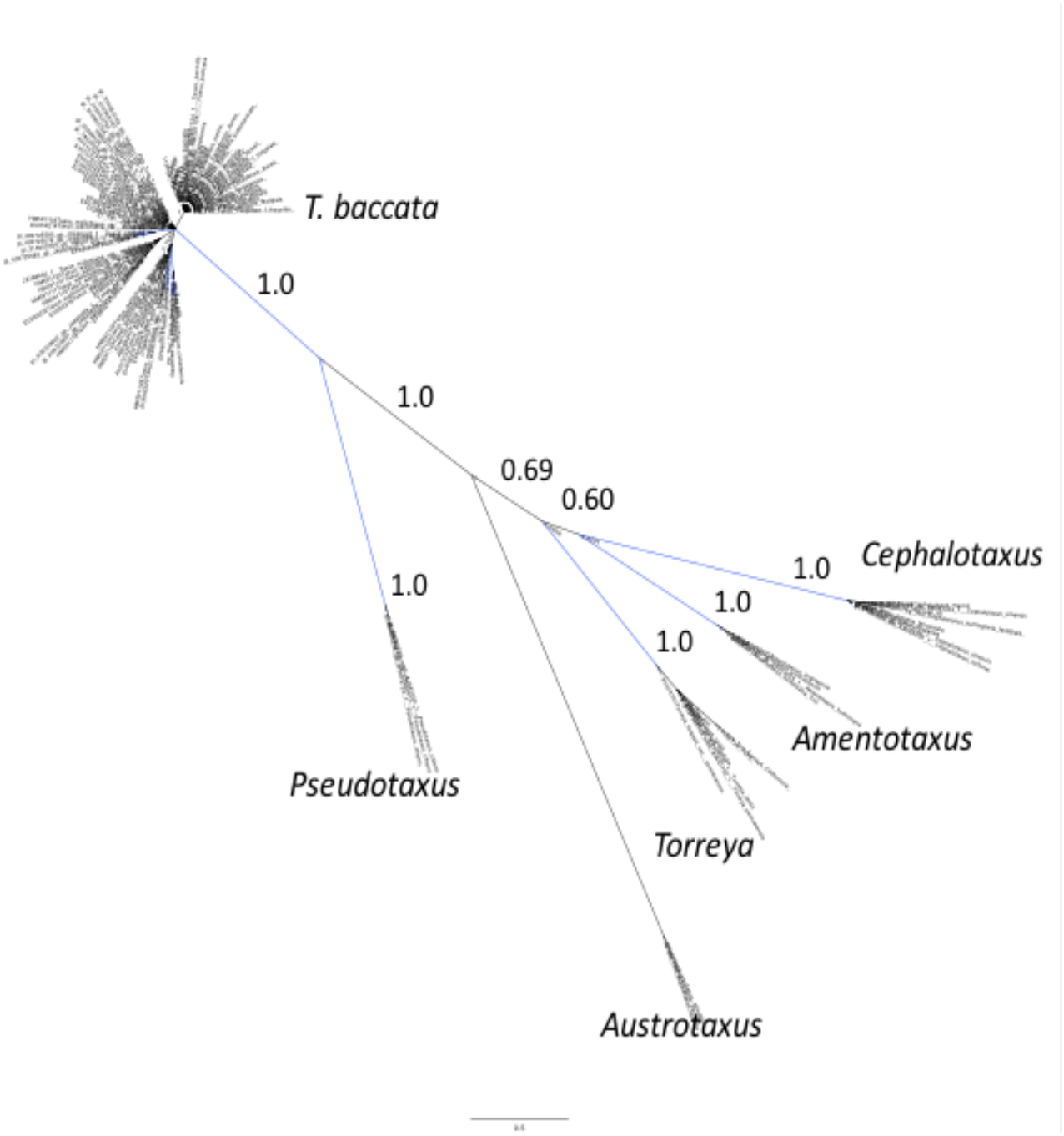


Figure 3.2. Phylogenetic reconstruction of Taxaceae based on trnL and trnL–F (trnL-F) plastid DNA. Bayesian posterior probabilities values shown above branches. Bootstrap values provided above or alongside branches.

In contrast, the phylogenetic analyses with nr*ITS* provide well-resolved and better-supported trees (Figure 3.3; Supplementary Figure 3.1). The *ITS* matrix with 119 species was 1150bp long with 333 segregating sites. A combined analysis was therefore undertaken with the combined *ITS, trnL* intron and *trnL-F* sequences using both ML and BI (86 sequences, matrix 1511bp long and 361 segregating sites). The BI tree is shown in Figure 3.3 with bootstrap values from the ML analysis given below the branches and BI

posterior probabilities above the branches. The combined tree shows that *Taxus brevifolia* is sister to *T.globosa* and *T.flordiana* (BS = 86; PP = 1). *Taxus fauna* groups with *T.contorta* (=*T.wallichiana* ssp. contorta) (BS = 99; PP = 1). The trees show that *Taxus brevifolia* is sister to *T. globosa* and *T. floridana* (BS=94; PP=1). *Taxus fuana* groups with *T. contorta*. *Taxus wallichiana* is resolved as monophyletic but its varieties *T. wallichiana* var. *mairei*, var. *chinensis* and var. *wallichiana* are not monophyletic although individuals within variety do generally group together and there is not firm support from PP or BS for the lack of monophyly for these varieties. The relationships of *T. baccata*, *T. canadensis, T. cuspidata* are not well resolved and there is no evidence for their monophyly.

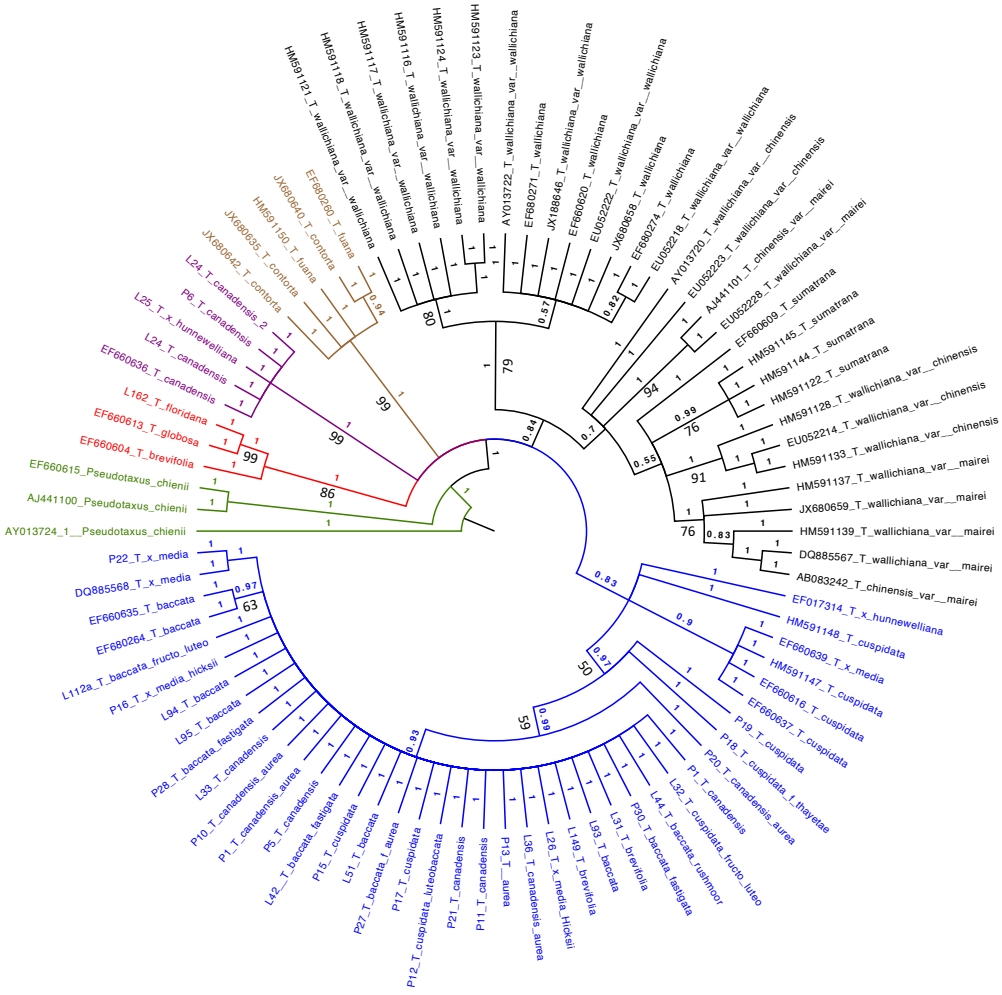


Figure 3.3. Bayesian inference tree of combined Taxus nrITS and plastid trnL and trnL-F sequences. Posterior probability values shown above the branches and maximum likelihood bootstrap analysis support the values shown below the branches. The relationship between T. baccata, T. canadensis and T. cuspidata are not well resolved. T. brevifolia is sister to T. globosa and T. flordiana. Taxus fauna groups with T. contorta. Taxus wallichiana is monophyletic and its varieties generally group together.

The TCS networks for both *ITS* (Figure 3.5) and *trnL-F* (Supplementary Figure 3.2) support the BI and ML analyses of these genes and again show the non-distinction of *T. baccata*, *T. canadensis, T. cuspidata.* Other network building methods Minimum spanning network and Median joining network (Bandelt et al., 1999, Clement et al., 2000, Clement et al., 2002) showed the same patterns (data not shown). The networks support the clear separation of *T. wallichiana* on one side of the network and an unresolved group of *T. baccata, T. canadensis* and *T. cuspidata* on the other side. The other species are positioned between these groups. The *ITS* sequence haplotypes are largely species specific except sequence 1 ,2, 11. Haplotype 2 includes *T. floridana* and *T. globosa*, haplotype 11 includes *T. cuspidata* and *T.* x*media* and haplotype 1 is the most common sequence type including representatives from *T. baccata*, *T. canadensis* and *T. cuspidata.*

The results of the anatomical study of leaf impressions are provided in Appendix 3.3 to 3.16. An example is shown in figure 3.4. Rows of stomata and papillae were observed and used in association with Spjut (2007b) key to identify the species. However, the results were inconclusive and hard to interpret (see Discussion section for more detail).

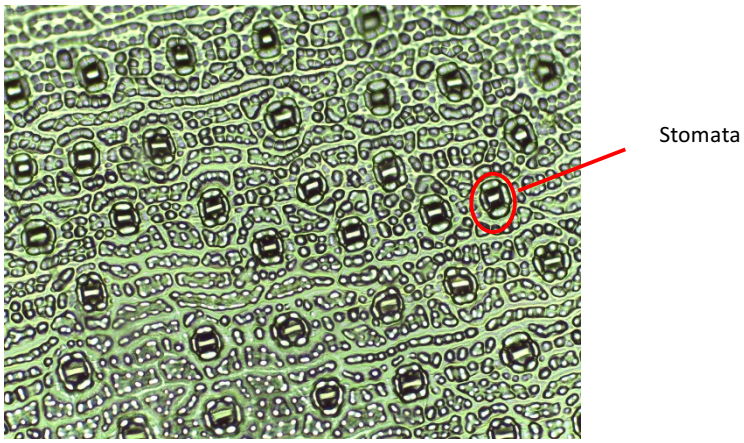


Figure 3.4. P16 T. x media "Hicksii" at 20x magnification showing rows of stomata.

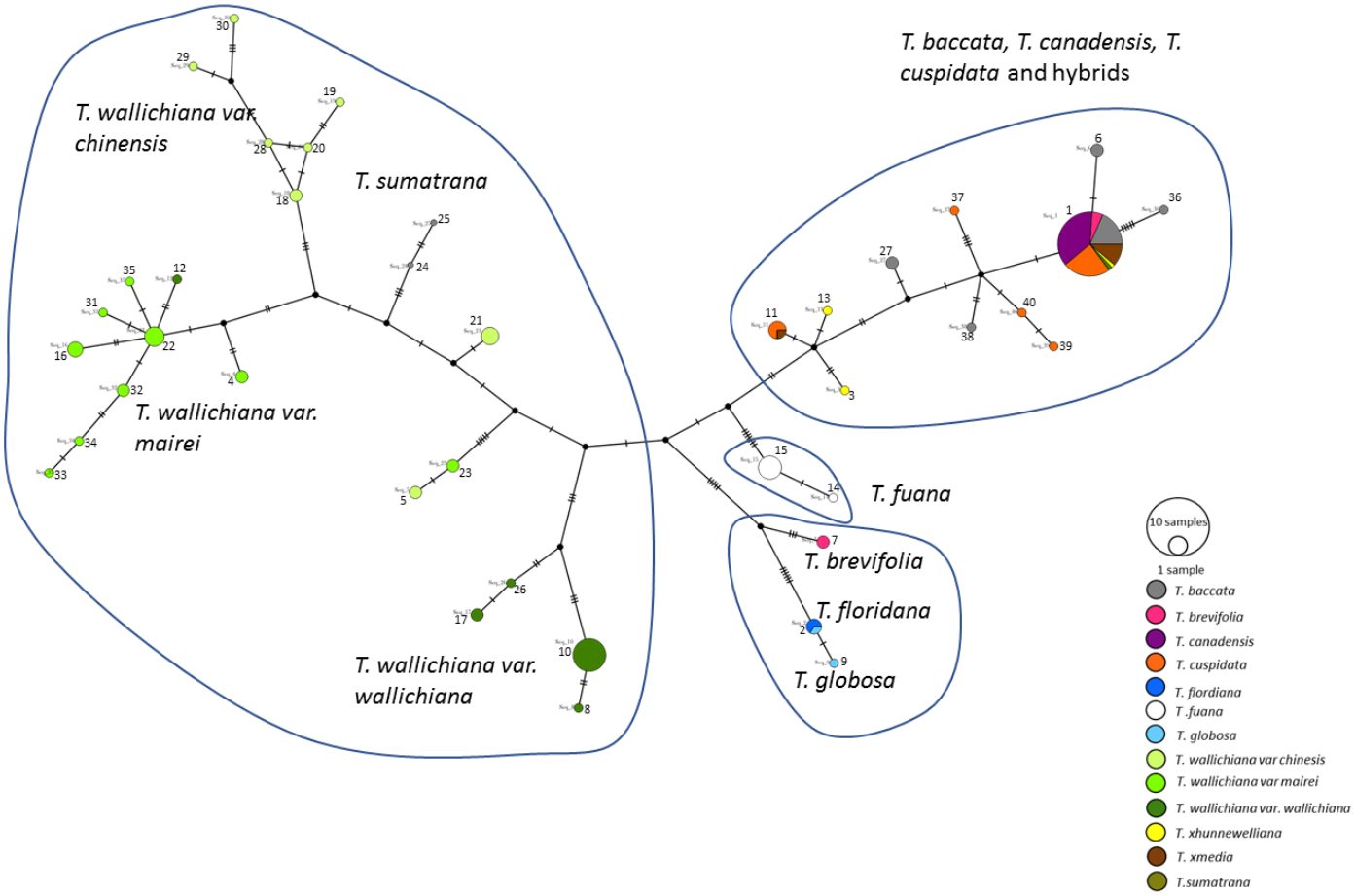


Figure 3.5. TCS network for ITS sequences. Network shows non-distinction of T. baccata, T. canadensis and T. cuspidata, whilst there is clear separation of T. wallichiana and a T. brevifolia, T. floridana and T. globosa group.

Sequence heterogeneity was detected in the uncloned *ITS* PCR products/sequences of *T.* x *media* and T. x *hunnewelliana*. Polymorphisms were detected at several sites and these can be mapped to the corresponding bases in their parental sequences (Figure 3.6). For example, *T.* x *media* has both a C and a G at position 142. This polymorphism is explained by the presence of the C in one putative parent (*T. cuspidata*) and G in the other parent (*T. baccata*). Similar polymorphisms can be seen for *T.* x *hunnewelliana*.

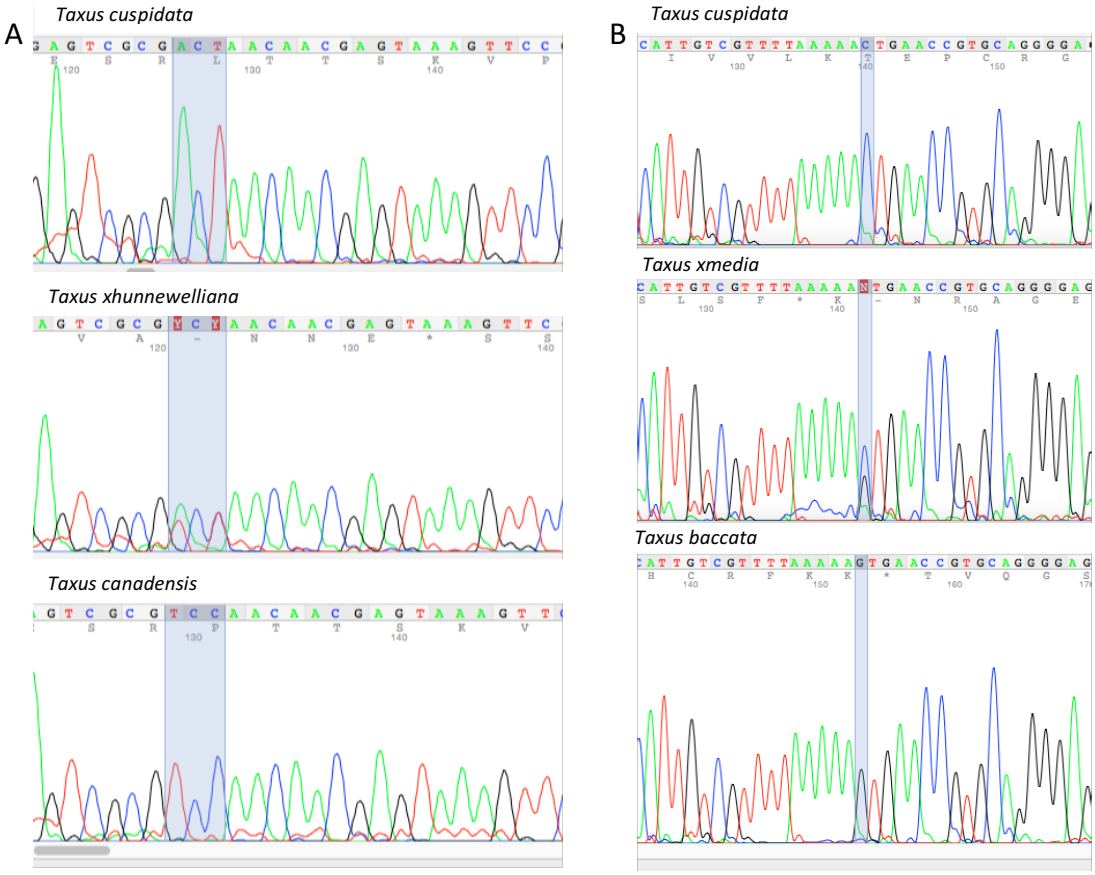


Figure 3.6. Sequence heterogeneity in T. x hunnewelliana (A) and T. x media (B) and corresponding sequences in their putative parental species.

3.5. Discussion

3.5.1. Taxaceae phylogeny

Our results show support for the recognition of six genera, *Amentotaxu*s, *Austrotaxus, Cephalotaxus*, *Pseudotaxus*, *Taxus* and *Torreya* in Taxaceae and each of the genera are clearly resolved as monophyletic (Figure 3.2; Supplementary Figure 3.1). This finding supports the taxonomic treatment of Christenhusz et al. (2011) who combined Cephalotaxaceae and Amentotaxaceae with Taxaceae to include 28 species in the same six genera. It is also consistent with the phylogenetic study of Price (2003) based on *rbcL* and *matK* that showed Taxaceae to be monophyletic when *Cephalotaxus* and *Amenotaxus* are included. Other authors have chosen to separate these genera into three families. For example, Hao et al. (2008b) used a phylogenetic analysis of the sequences of five chloroplast (*mat*K, *rbc*L, *trn*L, *trn*L-*trn*F spacer) and one nuclear molecular marker region (*ITS*), both individually and in combination, to support the division of the species into three allied families, Taxaceae, Cephalotaxaceae and Amentotaxaceae.

Pseudotaxus and *Taxus* are closely related sister genera with their only known morphological distinction being the difference in colour in the stomatal bands and aril. *Pseudotaxus* has white arils (Fu et al., 1999) and is native to south eastern China (north Fujian, north Guangdong, Guangxi, Huan, Jiangxi and Zhejiang) (Auders et al., 2012, Farjon, 2010a, Eckenwalder, 2009). The sister status of *Pseudotaxus* to *Taxus* is also strongly supported in this study but there is not enough support for the formulation of an infrageneric classification of Taxaceae. Elpe et al. (2018) conducted a phylogenetic study, using a combined dataset of two nuclear and two chloroplast gene regions. They used both maximum likelihood and Bayesian inference. The results of Elpe et al (2018) fully supported *Cephalotaxus* as a sister group to Taxaceae. Within Taxaceae two tribes are supported as monophyletic, Taxeae and Torreyeae. Taxeae consists of *Austrotaxus, Pseudotaxus* and *Taxus,* whilst Torreyeae comprises of *Amentotaxus* and *Torreya.* Ghimire and Heo (2014) adopting a cladistic approach to investigate the Taxaceae also found *Pseudotaxus* to be sister to *Taxus* with high bootstrap support.

3.5.2. Taxus phylogeny

The phylogenetic and network analyses reported here were based on the nuclear ribosomal *ITS* and plastid *trnL-F* regions which have been used in the past to investigate the phylogentics of the genus. Li et al. (2001) used sequences of the *ITS* regions to show genetic diversity among some *Taxus* species but their sampling was limited to one *Pseudotaxus chienii* sample and 14 *Taxus samples* (*T. baccata, T. brevifolia,* two *T. canadensis, T. chinensis, T. chinensis var mairei,* two *T.cuspidata, T. globosa, T. xhunnewelliana, T. xmedia, T. wallichiana*). Furthermore, Shah et al. (2008b) used sequences of *ITS* and *trnL-F* along with principal component analysis to determine the taxonomic and geographical boundaries between *Taxus* species but they only included *T. baccata, T. wallichiana* and *T. fuana.*

The phylogenetic analyses with nr*ITS* and combined analyses with plastid DNA reported here provides most resolution and support for groupings (Figures 3.3 and 3.4; Supplementary Figure 1) and indicates that *Taxus brevifolia* groups with *T. globosa* and *T. floridana* and that *Taxus fuana* groups with *T. contorta*. *Taxus wallichiana* is resolved as monophyletic but its varieties *T. wallichiana* var. *mairei*, var. *chinensis* and var. *wallichiana* are not monophyletic, although individuals within variety do generally group together. *Taxus baccata*, *T. canadensis, and T. cuspidata* are closely related but are not well resolved. There is little evidence for their monophyly except for one group of *T. canadensis* (in the *ITS* tree). The networks support the clear separation of *T. wallichiana* from an unresolved group of *T. baccata, T. canadensis* and *T. cuspidata* on the other side. The other species are found between these groups. The ITS sequences are largely species specific except sequence 1, 2, and 11. Sequence 11 includes *T. cuspidata* and *T.* x*media*, sequence 1 is the most common type including representatives from *T. baccata*, *T. canadensis*, *T. cuspidata and* sequence 2 includes *T. floridana* and *T. globosa*.

3.5.3. Species distinctiveness and identification

Some authors have argued that all *Taxus* should be combined into a single species (Pilger, 1903, Elwes and Henry, 1906, Dempsey and Hook, 2000, Dempsey, 2000) but our results,

combined with morphological and geographical evidence, support the division of *Taxus.* The distribution of these taxa is shown in Figure 3.7 including *T. baccata* (Europe, Canada), *T. brevifolia* (western North America), *T. canadensis* (eastern Canada, USA), *T. cuspidata* (Japan), *T. floridana* (South United States), *T. fauna (China)*, *T. globosa* (Mexico), *T. sumatrana* (China) and *T. wallichiana* (Eastern India). Comparison of the map in Figure 3.1 showing species distribution with the TCS network shows that it is possible that the groupings can partially be explained by historical biogeographical processes including continental drift (Figure 3.7).

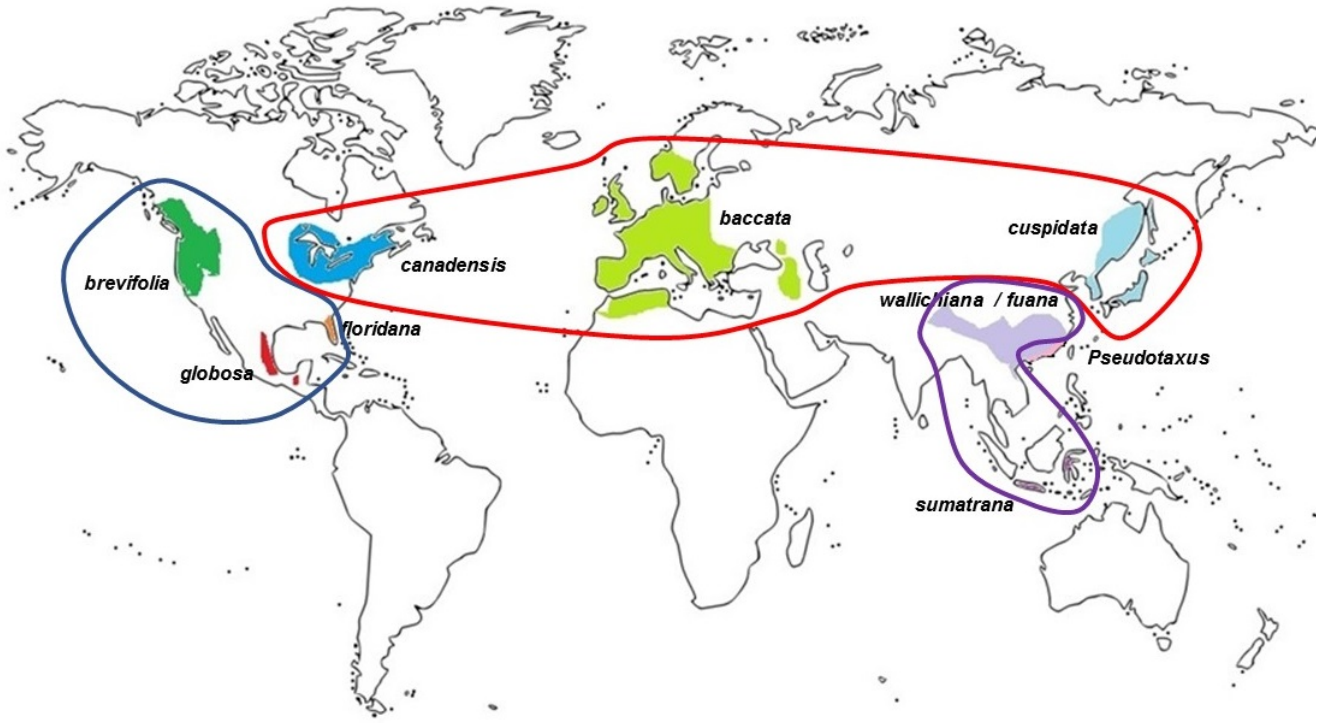


Figure 3.7. TCS network for ITS sequences in relation to the biogeography of Taxus. Blue, red and purple ellipses represent phylogenetic groupings from the TCS network.

The TCS network groups *T. baccata, T. canadensis, T. cuspidata* and hybrids together, and these grow in broadly similar biomes and latitudes. *Taxus brevifolia, T. floridana and T. globosa* are closely related in the TCS network and are all North American. One might expect *T. canadensis* to be in this group, however if we take continental drift into account it is not surprising that *T. canadensis* groups closely with *T. baccata* as their separation could have

been caused by vicariance as North America separated from Eurasia during the Cenozoic era. Similar patterns are known for other tree groups such as *Platanus* (Feng et al., 2005). *Taxus wallichiana* and *T. sumatrana* group together and are Asian and Indonesian, respectively. It is possible that they were formed via allopatric speciation from a common ancestor. E.g. *T. wallichiana* dispersing into Indonesia to become *T. sumatrana.* Alternatively, this pattern could be explained by long distance dispersal (e.g. via zoochory) and subsequent speciation. Female yews produce fruit which are consumed by birds and disperse the seed intact in their faeces (Snow and Snow, 2010). The fruits are eaten by several birds including winter flocking members of the thrush family as well as being hoarded and eaten by rodents (Williamson, 1978). However, Lavabre and García García (2015) did a study of the seed dispersal patterns of *T. baccata* across Spain and the results showed that the spatial distribution of the seeds in the landscape was heterogenous with the majority of the seeds consistently dispersed into forested microhabitats and almost none outside the forest. The results showed large scale consistency of the seed dispersal system of *T. baccata* and suggests that the generalized spatially restricted dispersal contributes to the lack of regeneration and absence of population expansion of the species all over its range. This could explain the separation between *T. brevifolia* and *T. canadensis* being separated geographically by the American prairies and *T. baccata* with *T. cuspidata* by the European and Russian steppes. A circumpolar distribution map of *Taxus* species from Hultén and Fries (1986) indicates interglacial records of *Taxus* on the European steppes. These fossil records could have been a link between *T. baccata* and *T. cuspidata.*

Taxus baccata, *T. canadensis* and *T. cuspidata* are problematic to differentiate using DNA sequence evidence alone but are sufficiently distinct in morphology and geographical distribution to merit species status. Other studies have provided evidence for the species status of several taxa. For example, Shah et al. (2008) strongly supported the distinctness of *T. baccata, T. wallichiana* and *T. fuana.* Spjut (2007a), Spjut (2007b) classified 24 species and 55 varieties into three groups according to differences in leaf epidermal and stomatal features recognizing 1) a *Wallichiana* group with subgroups *Wallichiana* and *chinensis,* 2) a

Baccata group with subgroups *baccata* and *cuspidata* and 3) a *Sumatrana* group (not divided). Our DNA studies and assessments of leaf anatomy (Supplementary Table 3.3; Supplementary Figures 3.3-3.16) showed support for the Group 1 and 2 of Spjut but not Group 3. Some of the problems with monophyly of species in the *ITS* tree could be explained by difficulties in taxon identification on the basis of morphology. Thus, a species identified *as T. canadensis* on the basis of morphology might actually be a *T. baccata* on the basis of DNA or anatomical evidence.

A subsample of the *Taxus* samples collected in the National Botanic Gardens, Glasnevin, Ireland, were identified to species *in situ* according to the key provided by Farjon (2010b). This was very difficult to do as the different species all look very similar, due to phenotypic plasticity. Also, a lot of the taxonomy can be subjective. For example, one of Farjon's characters for identification is the leaves on lateral branchlets arranged or not arranged in a V-formation. In the field this was very difficult to classify for some samples as the V-formation was not very convincing in some specimens. Spjut (2007a), Spjut (2007b) used leaf anatomical characters to assist in species identification. Samples of leaves were taken in our studies and epidermal cell patterns were visualized using cellulose acetate impressions according to the method of Sarvella et al. (1961). These impressions which were taken in the field were then observed in the laboratory under the magnification of 10X and 20X using a Leica ICC50 HD microscope. Rows of stomata and papillae were observed and used in association with Spjut (2007b) key to identify the species. The results were inconclusive (Supplementary Table 3.3) as Farjon and Spjut's keys lead to different results. For example, sample P14 which is labelled as *T. brevifolia* was keyed *in situ* to *T. cuspidata*, using Farjon (2010b). However, using Sput's key and the stomata impressions it was keyed to *T. recurvata,* which according to The Plant List (http://www.theplantlist.org) is a synonym of *T. baccata.* However, it is very easy to misinterpret Spjut's key as he uses phrases which could be easily misinterpreted such as "papillae are nearly medial" and leaves are "usually revolute". These are not distinct traits so could be interpreted differently from person to person. Also in some samples the midrib was mostly smooth with papillae only present in parts of the midrib. The leaf impression were carried out as a preliminary study, hence only

a small volume of samples were carried out and the results are presented in the supplementary table 3.3.

Elpe et al. (2018) developed a new identification key based on leaf anatomical characters using fluorescence microscopy. They found the presence of papillae on the abaxial midrib and on the adaxial leaf surface of *T. brevifolia,* to be a useful tool to separate *T. brevifolia, T. floridana, T. globosa* and *T. wallichiana* from other species. However, no differences were found between species which had a papillose midrib nor species which lacked this character. The key they produced does group *T. baccata* with *T. canadensis* and *T. cuspidata,* which further confirms our results. More data are needed for some species and this is important for conservation and utilisation. Therefore, we can conclude that a combination of both traditional taxonomy and DNA sequencing is the best way to identify the species.

A further complication is *ITS* copy repeat heterogeneity. Repeat units of nrDNA are typically homogenized by concerted evolution, so that only one predominant copy is present (Elder Jr and Turner, 1995, Hodkinson et al., 2002). This can cause complications in the interpretation of *ITS* sequences from closely related taxa. For example, in a hybrid line that has undergone subsequent cycles of sexual reproduction, the process of concerted evolution may homogenize copy types but sometimes favours one parental type over the other (Wendel, Schnabel, and Seelanen, 1995a). In the case of F1 hybrids concerted evolution could not have occurred by unequal crossing over, and two copy types, corresponding to the two parental species, might be detectable. However, some degree of concerted evolution may have occurred by gene conversion in non-F1 material. Nuclear DNA sequences, such as *ITS*, are also subject to recombination and, following a number of generations, individual repeats of the *ITS* sequence cannot only vary from each other but can also become highly heterogeneous themselves (Hodkinson *et al.* 2002). The repeat units can therefore become a mosaic of nucleotides from both parental types such that the original types are not easily distinguished (Wendel and Doyle, 1998). We found that many

ITS PCR products could not be sequenced without molecular cloning which supports the evidence for considerable sequence heterogeneity within some individuals. However, the heterogeneity also presented an opportunity to study hybridisation as outlined below, as different repeat types could be assigned to different parents of putative hybrid taxa.

3.5.4. Hybridisation

Sequence heterogeneity was detected in the uncloned *ITS* sequences of *T.* x *media* and T. x *hunnewelliana*. These putative hybrids had previously been determined from morphological studies. Polymorphisms were detected at several nucleotide sites and these can be mapped to the corresponding bases in their parental sequences (Figure 3.6) or cloned sequences from the same amplicon. This provides evidence in support of their hybrid status. *Taxus* x *hunnewelliana* is a hybrid of *T. cuspidata* x *T. canadensis* and *T. x media* is a hybrid of *T. baccata x T. cuspidata*. This finding supports other work especially by Collins et al. (2003) but needs to be interpreted carefully given the lack of resolution in the *T. baccata*, *T. canadensis* and *T. cuspidata* group. Some other work has examined hybridisation in the genus and found similar support for hybridisation. Collins et al. (2003) examined species distinction of *Taxus baccata*, *T. canadensis*, and *T. cuspidata* using Randomly Amplified Polymorphic DNA fingerprinting (RAPD) and DNA sequences and undertook an analysis of their reputed hybrids. They reported 16 polymorphic sites in the plastid *trnL-F* region that could be used for species identification and found three different plastid DNA haplotypes (Collins et al., 2003). They confirmed the hybrid origin and parentage of the *Taxus* hybrids (*T. cuspidata* x *T. canadensis* = *T. x hunnewelliana* and *T. baccata x T. cuspidata* = *T. x media*) using RAPD fingerprinting. Furthermore, 21 RAPD bands were species specific. Thus, given our findings and those of Collins et al. (2003) it seems likely that hybridisation and introgression is common among these three species and that they are not entirely distinct from each other.

3.6. **Conclusions**

This study supports the recognition of nine *Taxus* species (*T. baccata*, *T. brevifolia*, *T. canadensis*, *T. cuspidata*, *T. floridana*, *T. fauna*, *T. globosa*, *T. sumatrana* and *T. wallichiana*) and has determined some broad species groupings that are consistent with their historical biogeography. It has also found evidence of evolutionary reticulation that complicates the taxonomic understanding of the group. A broad phylogenetic framework of *Taxus* has therefore been provided to help guide comparative biological studies on the genus. For example, an accurate phylogeny is required to understand the evolution of the genes determining paclitaxel production, a chemical used for the treatment of ovarian, breast and lung cancer (McGuire et al., 1989, Holmes et al., 1991, Bristol Myers Squibb, 2011). The results are also important from a conservation perspective where clear taxonomic understanding is required (Elkassaby and Yanchuk, 1994, Kwit et al., 2004, Shah et al., 2008a, Mohapatra et al., 2009, Zhang and Ru, 2010).

3.7. **Acknowledgements**

We thank Colin Kelleher, National Botanic Garden, Glasnevin, Ireland for help with sampling and Mark Chase, Royal Botanic Gardens, Kew UK, for help with the DNA sequencing of some samples. The work was partly supported by Trinity College Dublin and Enterprise Ireland Basic and Strategic Grant SC/199/458.

Chapter 4: Phylogenetics of Paclitaxel biosynthesis genes in Taxus and allies

Phylogenetics of Paclitaxel biosynthesis genes in *Taxus* and allies

Coughlan P[1], Carolan JC[2], Hook IL[3], & Hodkinson TR[1]

[1]*Botany Building, School of Natural Sciences, Trinity College Dublin, the University of Dublin, Ireland*

[2] *Department of Biology, Maynooth University, Maynooth, Co. Kildare, Ireland*

[3]*School of Pharmacy and Pharmaceutical Sciences, Trinity College Dublin, the University of Dublin, Ireland*

4.1. Abstract

Paclitaxel is a diterpenoid produced by several species of *Taxus* and is used to treat various forms of cancer. It was originally extracted from the bark of the Pacific yew, *Taxus brevifolia* but high demand for Paclitaxel led the Pacific yew to become endangered. It is now semi-synthesised from the needles of *Taxus baccata.* We characterised two genes, Taxadiene synthase (*TS*) and 10-deacetyl baccatin III-10-0-acetyltranferase (*DBAT*), involved in the Paclitaxel biosynthesis pathway and compared variation in the genes among species of *Taxus.* All diterpenoids derive from a common substrate geranylgeranyl diphosphate (Toyomasu & Sassu 2010). The first dedicated step in taxol biosynthesis is the conversion of geranylgeranyl diphosphate to taxadiene by *TS* (Koepp et al 1995). *DBAT* catalyses the acetylation of 10-deacetyl baccatin III to baccatin III from which Paclitaxel is semi-synthesised. We developed new molecular primers to amplify and study the *TS* 1 to 5 exon regions. The primary aim was to assess molecular variation in the *TS* gene at DNA and protein levels and to test for evidence for selection on the gene. A sample of the *DBAT* gene was also sequenced in this study to look for evidence of molecular variation. The substitution rate is high in the *TS* gene with on average 43 changes per kbp and on average 22 changes per kbp in *DBAT.* Our tests with Tajima D gave large departures from neutrality for both *TS* and *DBAT* with the largest negative values occurring in the *TS* region, indicating strong natural selection has occurred at these loci. This evidence for selection is consistent with the findings from our phylogenetic analysis because the topology and groupings of the

gene trees from the *TS* and *DBAT* sequences are not consistent with those of our *ITS* and *trnL-F* trees. Our results show that there is variation in the sequence of *TS* and *DBAT* across species resulting in parallel changes in amino acids. The *TS* amino acid haplotypes determined are not species specific and there is considerable overlap among taxonomic groups.

Keywords

Biochemical pathway; Paclitaxel; Taxadiene synthase; *Taxus*; 10-deacetyl baccatin III-10-0-acetyltranferase.

4.2. Introduction

Species of the yew genus *Taxus* produce the secondary metabolite Paclitaxel (Taxol) which was first shown to have anticancer properties in the 1960's (Wani et al., 1971). It was first extracted by Dr. Monroe Wall from the bark of Pacific yew (*Taxus brevifolia*) but has since been identified in all *Taxus* species (Witherup et al., 1990, Vidensek et al., 1990, Mattina and MacEachern, 1994, ElSohly et al., 1997) and more recently in endophytic fungi associated with *Taxus* (Stierle et al., 1993, Flores-Bustamante et al., 2010, Hao et al., 2013). Bristol Myers Squibb (BMS) further developed Paclitaxel and gave it the trade-mark name Taxol. Paclitaxel is found in all parts of the yew tree except for the fleshy aril. If it is ingested, it can cause death by heart failure. Paclitaxel has been shown to be embryotoxic and foetotoxic in rabbits, and to decrease fertility in rats (Irish Medicines Board, 2007). Despite this toxicity, Taxol is used to treat ovarian, breast and lung cancer (Bristol Myers Squibb, 2011). Initially Taxol was synthesised from the bark of yew trees with 30 tons of bark required to produce up to 100 grams of purified Taxol. Stripping the bark kills the tree. This had consequences for the conservation of yew trees as the species became endangered. To contend with this unsustainable harvesting, BMS developed novel methods to extract Taxol from the needles of the tree instead (Witherup et al., 1990, Goodman and Walsh, 2001). The harvesting of needles does not kill the yew tree so makes it a renewable source. Hook et al. (1999) investigated seasonal variations of paclitaxel and 10-deacetylbaccatin III in

shoots of European yew collected from two locations, Ireland and France. No conclusion could be drawn as to the optimum time of year for harvesting, since this varies from tree to tree, depending on *T. baccata* variety, location and taxoid type (Hook et al., 1999). Taxoid is a member of a class of anticancer drugs derived from yew tree products.

It is not known how Paclitaxel synthesis *of Taxus* differs among species/cultivars, that can produce or sequester higher levels of the phytochemical than others, and whether the factors that govern these differences are environmental/genetic or a combination of both. Although some knowledge has been generated about the phylogenetic relationships within *Taxus* (Li et al., 2001, Kilmartin, 2002, Hao et al., 2008a), very little is known about molecular DNA variation in the genes involved in Paclitaxel production in the genus. This chapter adds to the existing knowledge of Paclitaxel by developing molecular primers to amplify and study the region of genes involved in the Taxol biosynthetic pathway and takes a molecular evolution approach to study genetic variation in these genes that might be important to variation in Paclitaxel production.

Considerable variation exists in Paclitaxel production in yew (Kelsey and Vance, 1992, Griffin and Hook, 1996, ElSohly et al., 1997, Hook et al., 1999, Poupat et al., 2000, Dempsey and Hook, 2000, Dempsey et al., 2003) but little is known about the control of this at a genetic level and hence the key factors influencing Paclitaxel production (Hao et al. 2009). Therefore, the focus of this study is on molecular variation in the taxadiene synthase (*TS*) gene involved in the production of Paclitaxel. Hao et al. (2009) examined patterns of evolution of the paclitaxel biosynthetic enzymes within *Taxus* and their results suggested that positive selection is driving divergence of *TS* in closely related *Taxus* species. We expand this study to include a greater diversity of species and assess genetic diversity of *TS* in the context of phylogenetic history inferred by our previous studies of the genus based on plastid and nrDNA (Chapter 3). Taxadiene synthase was chosen because of the availability of primers for its PCR amplification and sequencing and because it is the first major step of Taxol biosynthesis (Figure 4.1). It catalyzes the cyclisation of geranylgeranly

disphosphate to taxa-4(5), 11 (12)-diene (Koepp et al., 1995). It is a large gene with 13 exons and is approximately 4kbp in length (Hao et al. 2009, Chapter 1, this thesis). We assess differences within and between species and also among known hybrids and their known parents. In doing so, we used primers from the literature (Table 4.1) except for TS1(F), TS3(R), TS3(F) and TS5(R) that we newly designed from a *TS* sequence alignment file for a number of *Taxus* species to yield PCR products of approximately 600 to 700 base pairs when used together or in combination with published *TS* primers. The primary aim was to assess molecular variation at the DNA and protein level and to test for evidence for selection on the gene. In addition, some samples were sequenced for 10-deacetyl baccatin III-10-0-acetyltranferase (*DBAT*). *DBAT* catalyses the acetylation of 10-deacetyl baccatin III to baccatin III (Walker and Croteau, 2000). The full length of cDNA of *DBAT* is approximately 1323 base pairs over two intervals 1-435 and 659-1,546.

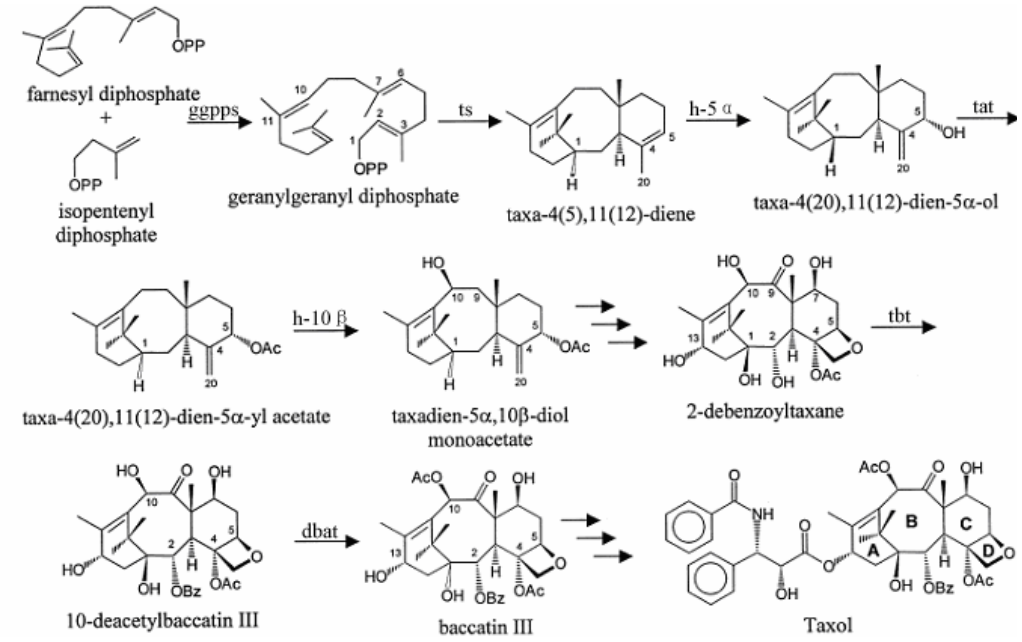


Figure 4.1. Taxol biosynthetic pathway and the genes involved. ggpps: geranylgeranyl diphosphate synthase; ts: taxadiene synthase; h-5_: cytochrome P450 taxadiene 5_-hydroxylase; tat: taxa-4(20), 11(12)-dien-5a-ol-O-acetyltransferase; h-10_: cytochrome P450 taxane 10_-hydroxylase; tbt: taxane 2a-O-benzoyltransferase; dbat: 10-deacetyl baccatin III-10-O-acetyltransferase. Multiple arrows indicate several as yet undefined steps. Taken from Guo et al., 2006. Further details of the Taxol biosynthesis pathway can be found in Walker and Croteau (2001), Jennewein et al. (2004), Croteau et al. (2006), Nims et al. (2006) and Guo et al. (2006).

4.3. Materials and Methods

4.3.1. Specimens

Fresh plant material was obtained from the National Botanic Garden, Glasnevin, Ireland. DNA samples were also available from the Trinity College Dublin (TCD) DNA Bank. Voucher specimens were kept for each sample, dried and stored in the TCD Herbarium. Some ground and dried leaf samples were also available from our previous work on the genus (Table 3.1).

4.3.2. Other Taxaceae genera

The *TS* gene was also amplified and sequenced from some related genera in Taxaceae including *Austrotaxus*, *Cephalotaxus*, *Podocarpus* and *Torreya*. *Cephalotaxus* was sister to *Taxus* in (Hao et al., 2008b) and Taxaceae comprises *Amentotaxus, Taxus* and *Torreya* (Farjon, 1998). *Cephalotaxus* is sometimes included in Taxaceae (Li and Fu, 1997). *Podocarpus* belongs to Podocarpaceae (Farjon, 2010b). Price (2003) made sequence comparisons of the chloroplast genes *rbc*L and *mat*K and suggested that the five genera of traditional Taxaceae plus *Cephalotaxus* could be treated as a single family. Within the expanded Taxaceae, three subfamilial lineages were found to be well supported by sequence comparisons: *Cephalotaxus; Amentotaxus* plus *Torreya*; and *Austrotaxus* plus *Pseudotaxus* and *Taxus*. Data by Coughlan et al. (in prep.; Chapter 3) strongly support the monophyly of these genera, place *Pseudotaxus* sister to *Taxus* and support a Taxaceae including *Amentotaxus, Austrotaxus, Cephalotaxus, Pseudotaxus, Torreya*, and *Taxus* (supporting the treatment of Price, 2003). Christenhusz et al. (2011) also combined these six genera (including 28 species) into Taxaceae.

4.3.3. DNA extraction and sequencing

DNA was extracted from 0.05-0.075g ground material or 0.075-0.1g of fresh material using a modified hot CTAB method of (Doyle and Doyle, 1987, Gawel and Jarret, 1991, Hodkinson et al., 2007). It was precipitated using 100% isopropanol, pelleted and washed with 70% ethanol and purified using JETquick Spin Columns (GENOMED Gmbh, Lohne, Germany).

DNA was then stored in TE buffer (10mM Tris-HCL, 1mM EDTA, pH8.0) at -80°C until required.

For amplification and sequencing of the exon 1-4 *TS* region, the forward primer TS1-4 (F) 5'-ATGGCTCAGCTCTCATTTAATGC-3' (Hao et al., 2009) and the reverse primer TS1-4 (R) 5'-CGCAGCCGCCGAATTTGTCCA-3' (Hao et al., 2009) were used (Figure 4.2, Table 4.1). These primers amplified an approx. 1.3 kbp fragment which is very large for direct Sanger sequencing, so internal primer combinations were developed here and used for sequencing and PCR amplification when necessary. These were as follows: TS1F 5'-CATGGCGATCTGTGGCACC-3', TS3R 5'-CACTCCTCTGTAAGCCTGG-3', TS3F 5'-CACAGCCAAGTAGAACAAGG-3', and TS5R 5'-CAACCAGCGAAAGGCGTTCC-3'. These new primers were manually designed from a *TS* sequence alignment file for a number of *Taxus* species obtained from GenBank using BioEdit (Hall, 1999) to yield PCR products of approximately 600 to 700 bps when used together or in combination with published *TS* primers (Figure 4.1). For amplification and sequencing of the *DBAT* gene region, the forward primer 5'-ATGGCAGGCTCAACAGAATTTG-'3 (Hao et al., 2009) and the reverse primer 5'-TCAAGGTTTAGTTACATATTTGTTTG-'3 (Hao et al., 2009) were used (see Figure 1.2 in Chapter 1).

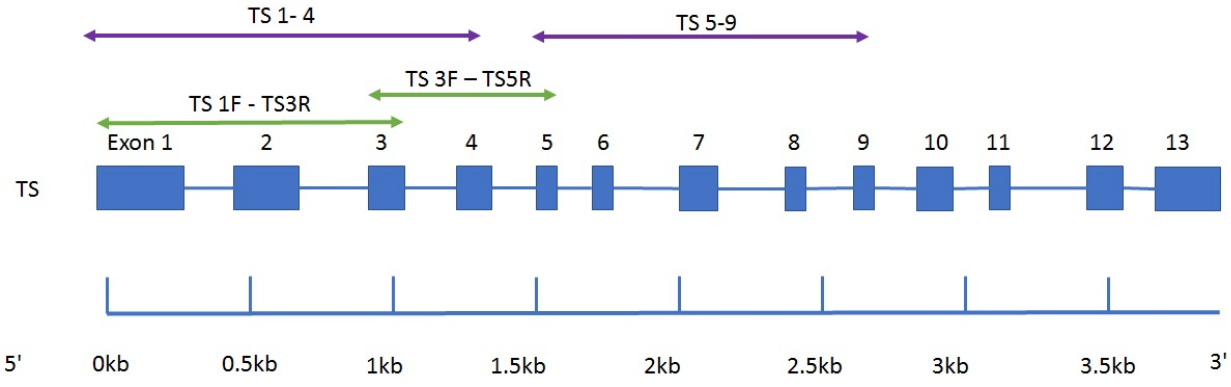


Figure 4.2. TS gene region showing 13 exons and the position of the primer sites used for PCR amplification and sequencing.

Table 4.1. Primers used for amplification of TS and DBAT gene.

Gene	Primer Code	Primer Sequence	Reference
TS	TS1-4(F)	5'-ATGGCTCAGCTCTCATTTAATGC-3'	Hao et al., 2009
TS	TS1-4(R)	5'-CGCAGCCGCCGAATTTGTCCA-3'	Hao et al., 2009
TS	TS5-9(F)	5'-TGGACAAATTCGGCGGCTGCG-3'	Hao et al., 2009
TS	TS1(F)	5'-CATGGCGATCTGTGGCACC-3'	Designed by James Carolan*
TS	TS3(R)	5'-CACTCACTCTGTAAGCCTGG-3'	Designed by James Carolan*
TS	TS3(F)	5'-CACAGCCAAGTAGAACAAGG-3'	Designed by James Carolan*
TS	TS5(R)	5'-CAACCAGCGAAAGGCGTTCC-3'	Designed by James Carolan*
DBAT	DBAT(F)	5'-ATGGCAGGCTCAACAGAATTTG-'3	Hao et al., 2009
DBAT	DBAT(R)	5'-TCAAGGTTTAGTTACATATTTGTTTG-'3	Hao et al., 2009

*Designed by James Carolan but tested and developed further here.

PCRs were carried out in 12.5µl reactions using BIOLINE Biomix (Bioline Reagents, UK). The *TS* PCR amplifications were prepared using 5.375µl ultrapure water, 6.25 µl Biomix, 0.125 µl (4pmol) forward and reverse primers and 0.625µl of column cleaned total DNA (ca. 100ngµl^{-l}). The reaction conditions for *TS* were as follows: denaturation at 95°C for 1min followed by 38 cycles of 30sec at 95°, 1min at 59°C, 1min 30sec at 72°C and a final extension at 72°C for 1min in an Applied Biosystems Verti 96 well thermal cycler (alternative annealing temperatures of 50°C and 55°C were used for *Torreya* and *Podocarpus* respectively). For 1F3R and 3F5R internal primers the annealing temperature used was 55°C and 0.125 of 20pmolµl^{-l} primer was used. The reaction conditions for *DBAT* were the same as *TS* but with an annealing temperature of 49°C. 3µl PCR products were run on a 1.2% agarose gel to check for amplification. Successfully amplified DNA fragments were purified using the ExoSap Method. 0.3µl of exonuclease, 2µl of shrimp alkaline phosphatase (1U/µl) and 2.7µl of sterile ultrapure water was added to 5µl of PCR product and incubated at 37° for 30mins followed by 82° for 20mins. Purified PCR products were then sequenced on Applied Biosystems 3130xl genetic sequencer according to the manufacturers' protocol with the same primers used for initial amplification.

4.3.4. DNA cloning

Some samples required cloning because the PCR product was heterogeneous. Cloning was performed using a Thermo Scientific CloneJET PCR cloning kit (Fermentas, Lithuania). The PCR product was inserted into the pJET1.2/blunt cloning vector which was then transformed into *E. coli* cells. The cells were then incubated and grown overnight at 37^{o}C. Eight single colonies where chosen randomly from the agar plate and a PCR was performed on each colony using the same primers as the initial pre-cloning amplification. A small part of the colony was picked directly from the agar and placed directly in the reaction using a sterile pipette tip. The parameters for the PCR were the same as above with the exception that there was an additional 10mins pre-melt at 94°C and 10mins final extension at 72°C. The cloned PCR products were purified using JETquick spin columns and sequenced as described above.

4.3.5. Sequence analysis, annotation and translation

To obtain a contiguous sequence for the target DNA region, forward and reverse sequence reads were assembled in Geneious Pro 5.6.4. (Biomatters Ltd.). The sequences were aligned in Geneious using highest sensitivity, using either Muscle or Geneious algorithms with default settings. Alignments were translated into amino acid sequences and exons aligned by codons.

The sequences were then manually aligned if necessary. Synonymous and non-synonymous mutations were tabulated and compared across the gene among exons (in Geneious) and assessments made of nucleotide diversity in MEGA 7 (Kumar et al., 2016). Tests were conducted in MEGA 7 (Kumar et al., 2016) to assess non-neutrality and evidence of selection and adaptive evolution. Tajima's D test of neutrality was used to test for overall non-neutral variation.

4.3.6. Phylogenetic analysis

The aligned matrix was imported into MEGA 7 (Kumar et al., 2016) for Maximum Likelihood (ML) analyses. Gaps in introns smaller than 10bp were coded as missing data, unless they

were found in regions where there was an obvious tandemly arranged duplication in one sequence that was clearly due to a single mutation (a duplication). Such duplications were scored as only one character in the subsequent phylogenetic analyses. Gaps larger than 10bp were excluded from the analyses.

For the ML analyses, the best fit substitution model was determined by the Model Selection function in MEGA 7 (Kumar et al., 2016) and was found to be the Tamura 3-parameter model (T92) with gamma distributed rate heterogeneity and estimated proportion of invariant sites (G+I). ML was performed in MEGA7 with 1,000 replicates of random sequence addition and nearest neighbour interchange (NNI) branch swapping. Initial tree(s) for the heuristic search were obtained automatically by applying Neighbor-Join and BioNJ algorithms to a matrix of pairwise distances estimated using the Maximum Composite Likelihood (MCL) approach, and then selecting the topology with superior log likelihood value. Bootstrap support was calculated from 1,000 replicates with the same settings as the initial search. Analyses were conducted using 1) the whole gene, and 2) on a partition of the dataset that included third position (mostly silent mutations) and non-coding introns because these sites are less likely to be influenced by selection and hence more accurately reflect phylogenetic history of the taxa following (Christin et al., 2007). Haplotype networks were constructed using PopArt (Leigh and Bryant, 2015). Sequences were coded according to taxon name and haplotype networks constructed using TCS following Clement et al. (2002).

4.4. **Results**

A total of 34 individual new *TS* sequences and 32 samples from GenBank were combined into a matrix of 66 sequences for comparison. We sequenced some outgroup taxa including *Cephalotaxus, Podocarpus, Torreya* and *Austrotaxus* but these were not complete or were hard to align with confidence and are not shown in Table 4.2. This demonstrates the presence of the gene in the related genera and shows the transferability of the primers. The

newly designed primers presented here were also useful for improving amplification and sequencing of regions within exons 1-4.

It can be seen from the summary table of molecular variation that the substitution rate is high in the *TS* gene with on average 43 changes per kbp (Table 4.2). Synonymous mutations were common (53 in total; range among exons = 0-10). Replacement mutations were more common (total number of non-synonymous mutations = 58; range among exons = 0-15). Variable sites in the amino acid translation for exons 1-4 are shown in Figure 4.3. This is reflected in the TCS network built on nucleotide codon data (Figure 4.5; exons only tree) of the entire gene that show some phylogenetic pattern. A different phylogeny is found when ML trees were reconstructed using third position and introns only compared to trees constructed with all codon positions (Figure Supplementary 4.1). This is closer to the combined nr*ITS* and plastid *trnL* and *trnL-F* tree from Chapter 3.

Molecular variation in the *DBAT* gene is also summarised in Table 4.2. There are 11 variable sites of which 5 are non-synonymous (Table 4.2; Figure 4.4) and 6 synonymous (Table 4.2) in the 439 bp of the gene sequenced. There were 25 substitutions per 1kb and a nucleotide diversity ($\boldsymbol{\pi}$) of 0.0058 which is higher than the average exon nucleotide diversity for *TS* (0.0028) but within the range of variation detected for its exons individually. There does not appear to be any taxonomic pattern in the mutations.

Table 4.2. Molecular variation of TS and DBAT genes. TS is partitioned by exon

Exon	n	bp	Syn	Non-Syn	S	**π**	Substitutions/1kbp	Tajima *D*
TS								
1	66	324	4	7	11	0.001483	33.95	-2.200
2	66	313	10	15	25	0.013244	79.87	-0.663
3	66	178	7	6	13	0.006351	73.03	-1.680
4	66	215	2	14	16	0.005269	74.42	-1.966
5	66	121	4	2	6	0.000445	49.59	-2.317
6	66	113	1	4	5	0.000487	44.25	-2.173
7	66	215	1	5	6	0.000256	27.91	-2.308
8	66	113	7	1	8	0.000982	70.80	-2.423
9	66	99	1	1	2	0.000132	20.20	-1.621
10	66	210	5	2	7	0.000422	33.33	-2.361
11	66	139	0	0	0	0	0	0
12	66	249	5	0	5	0.000337	20.08	-2.110
13	66	300	6	1	7	0.000227	23.33	-2.396
Total	66	2589	53	58	111	0.002829	42.87	-2.374
DBAT								
	n	bp	Syn	Non-Syn	S	**π**	Substitutions/1kbp	Tajima *D*
DBAT 1	*10*	*439*	*6*	*5*	*11*	*0.005821*	*25.06*	*-1.535*
DBAT 2	*21*	*261*	*3*	*2*	*5*	*0.002518*	*19.16*	*-1.589*

n=number of sequences; bp=length of exon; Syn=number of synonymous substitutions; Non-Syn=number of non-synonymous substitutions; S=number of segregating sites; **π**=nucleotide diversity

Note: GenBank sample AJ320538 was entered in a replicate of 4 times. GenBank sample DQ305407 was entered twice. *DBAT* 2 is a different matrix of *DBAT* with more taxa (n) but shorter sequence (only a subset of *DBAT* 1).

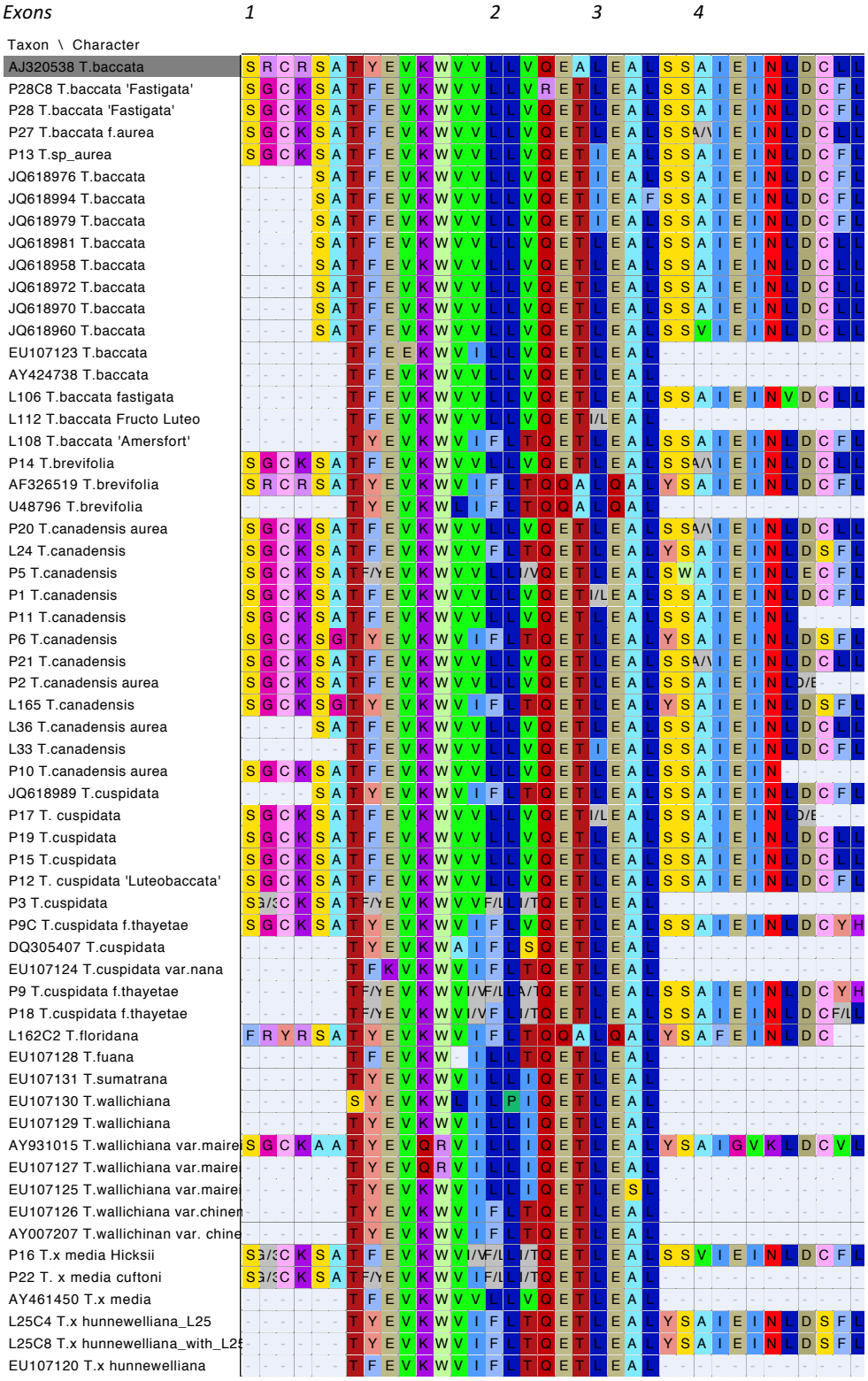


Figure 4.3. Variable amino acid sites in the TS gene (exons 1-4). Gaps are unsequenced regions in those samples. Letters represent amino acid code see supplementary table 4.1.

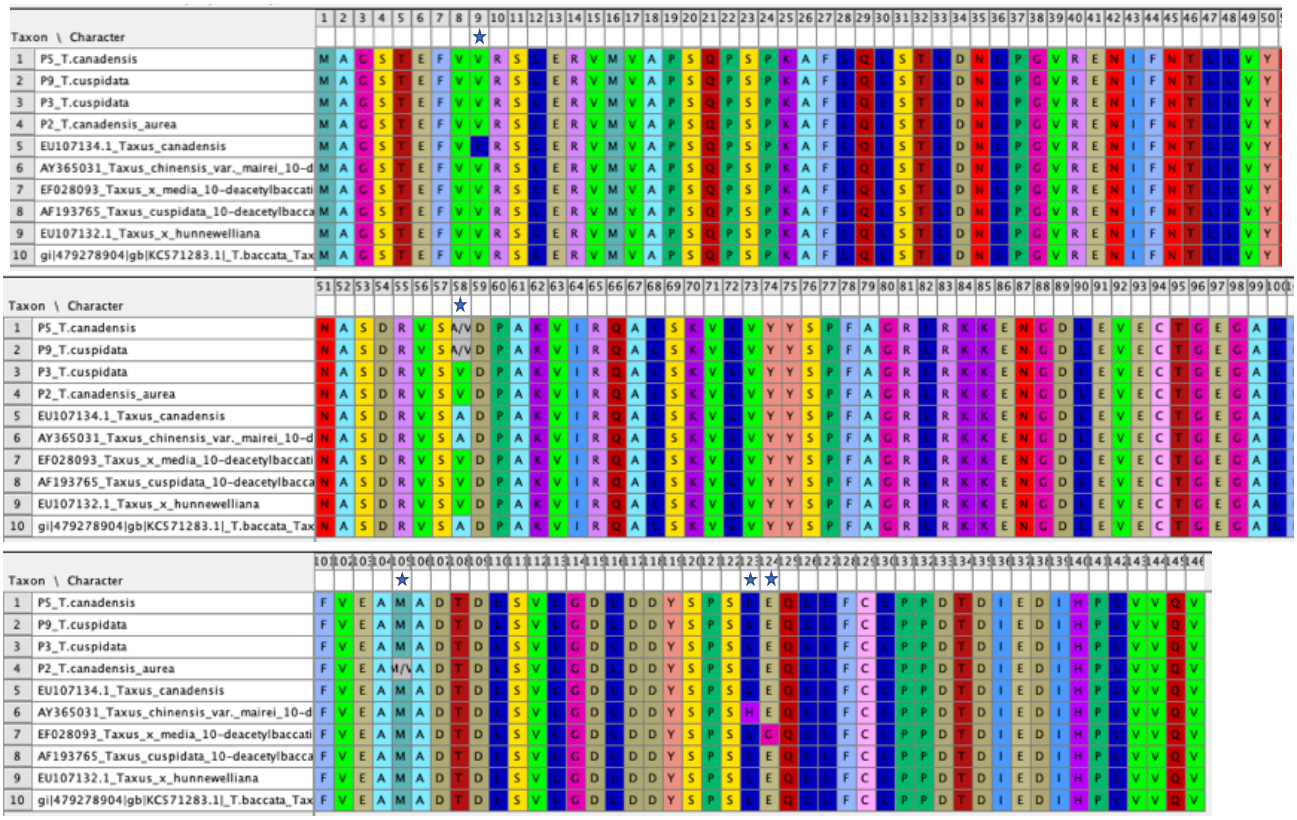


Figure 4.4. Variable amino acid sites of DBAT. Variable sites highlighted with star.

4.4.1. Selection and adaptive evolution

The Tajima's D statistic showed -2.374 across the whole gene and a range from 0 to -2.423 across the exon partitions of the gene. The Tajima D statistic for DBAT gave a lower value of ca. -1.5. Phylogenetic trees based on the entire *TS* gene did not determine the monophyly of *Taxus* species or reflect the phylogenetic relationships of taxa inferred by nuclear ribosomal *ITS* and other genes (Chapter 3). We therefore, repeated phylogenetic analysis using 3rd position and intron sites only (Figure Supplementary 4.1) and a TCS network (Figure 4.5). They show some phylogenetic pattern but do not resolve relationships as well as our analyses with plastid *trnL-F* and nuclear *ITS* (Chapter 3). The failure of the whole *TS* gene to reflect phylogenetic groupings may be an indication of positive selection (like the Tajima D test) and we would expect improvement in phylogenetic reconstruction with the inclusion of neutral third position and intron sites only. We saw some improvement which is consistent with such expectations.

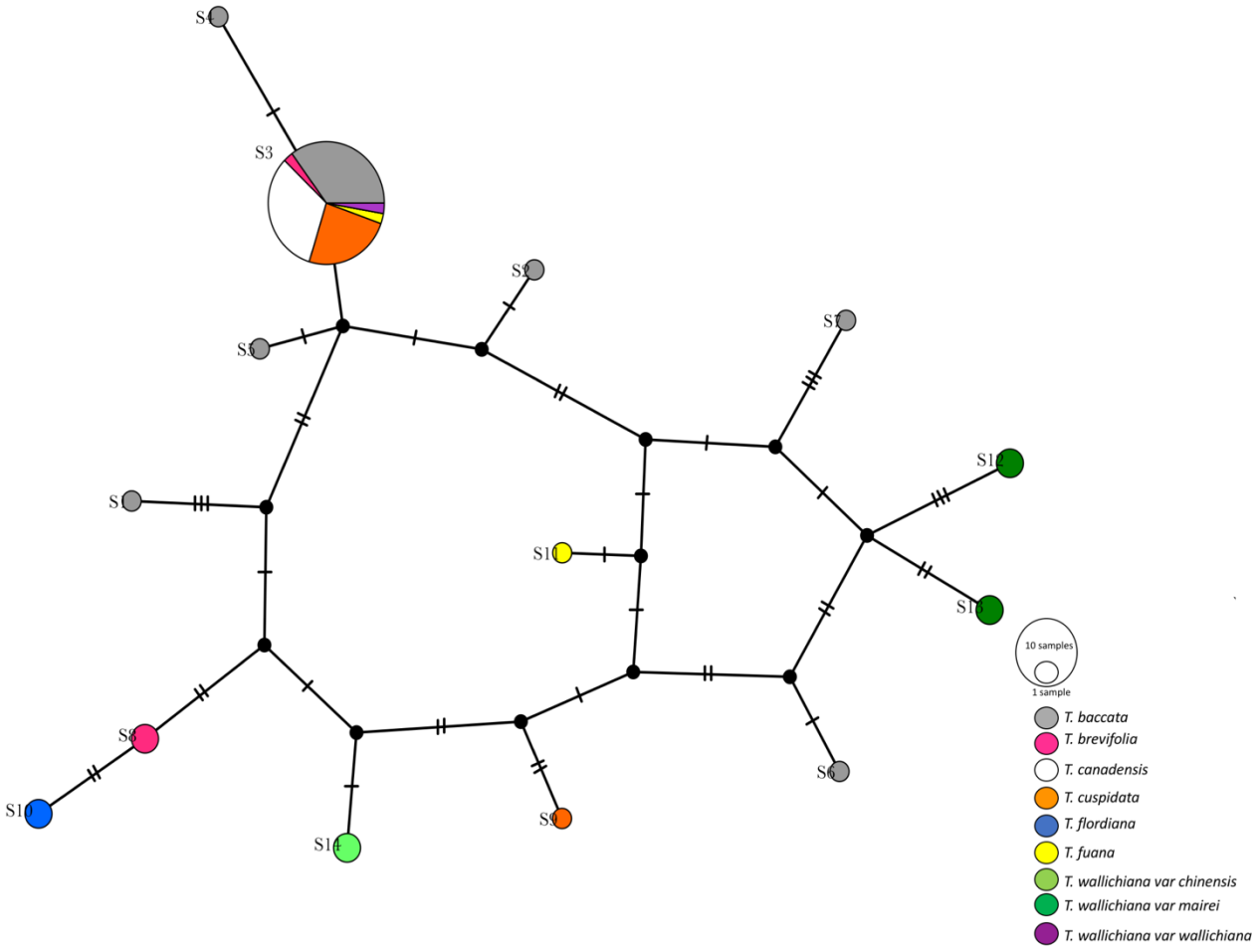


Figure 4.5. TCS network of TS gene variation (all codon sites).

4.5. **Discussion**

Our analysis consisted of 66 individuals which provides additional insight of *TS* gene sequence variability to that of previous studies such as Hao et al. (2009), which included only 19 individuals. In addition, we sequenced *Taxus floridana* for which no published or publicly available sequence for the *TS* gene before has been reported. Although van Rozendaal et al. (2000) did not identify any taxanes in Taxaceae genera other than *Taxus*, the *TS* gene was present and amplifiable in *Austrotaxus, Podocarpus* and *Torreya* which is a novel finding. However, other genes in the biochemical pathway may be missing/non-functional so the production of taxanes although possible can only be

demonstrated by phytochemical analysis. We also sequenced *Cephalotaxus TS* for which one previous sequence is available on GenBank.

Using the primers of Hao et al. (2009) several difficulties with PCR amplification and sequencing were encountered. To circumvent this, novel primers for the *TS* gene were designed that allowed the successful amplification of exons 1-4. Future work is required to develop more primers to cover *TS* exons 5-13. Hao et al. (2009) only sequenced and studied *TS* exons 1-9. They newly generated 12 *TS*, 14 *DBAT* and 16 *ITS* sequences.

4.5.1. Molecular variability of TS and DBAT

Our results show that there is considerable variation with the *TS* gene and some exons are more variable than others. On average there are 43 substitutions per kbp (Table 2). Replacement mutations were more common (total number of non-synonymous mutations = 58) than synonymous mutations but these were also common (53 in total). Variable sites in the amino acid translation for exons 1-4 are shown in Figure 4.3. There is some conservation of haplotypes within species but also considerable parallel change among species and taxa. Therefore, amino acid haplotypes are not species specific and there is considerable overlap among taxonomic groups. Table 4.2 shows that *TS* Exon 11 has no variation. It is most likely a highly conserved gene where mutations are not tolerated. Trapp and Croteau (2001) found that the *TS* exons 10-13 are highly conserved in organisation and catalytic function. Future work should increase sampling around exon 11. Exon 2 has the highest substitution rate per base pair, but we would expect more diversity in exons 1-4 because there are more species sampled for those exons. Exon 1 has quite low variability meaning that the start of the gene is also conserved relative to the others.

Synonymous and non-synonymous variation was determined for both *TS* exon 1-4 and *DBAT* exon 1. There are numerous non-synonymous mutations in the *TS* gene which result in variability in the amino acid sequence (Figure 4.3) and also in *DBAT* (Figure 4.4). It is possible that positive selection has occurred to make the yew tree more toxic to herbivorous animals or pathogens. The minimal toxic dose of taxines varies between species of animals. For example, horses are a lot more sensitive to yew poisoning than

chickens. The lethal minimum dose per kg of body weight is 16.5g of yew leaves for chickens and 0.2-0.4g for a horse (Wilson et al., 2001). The Tajima's *D* statistic can be used for detecting natural selection and selective sweeps in particular, i.e. the effect of an advantageous mutation going to fixation in the population (Korneliussen et al., 2013). It is most often used at population level (population genetics) but it can be applied at species level when divergence is low (Nei and Kumar, 2000). It can also be caused by recent population expansion. After fixation of an advantageous allele, there is an excess of rare alleles in the population. This will cause Tajima's *D* to become negative. Our tests with Tajima D gave large departures from 0 (=neutrality) for both *TS* and *DBAT* with the largest negative values occurring in the *TS* region, indicating strong natural selection has occurred at these loci.

Future research should examine the non-neutrality aspects of *TS* and *DBAT* evolution in more detail, as our results indicate that selective sweeps could have occurred. For example, there are many other ways to test for selection and identify codons under selection such as using dN/dS ratio tests where a dN/dS less than 1 can suggest purifying selection; a ratio of 1.0 as neutral and a ratio of greater than one as evidence of positive selection (Nei and Kumar, 2000). For this, the evolutionary pathway method of Nei-Gojobori (Nei and Gojobori, 1986) can be used to compare codons. To test for the overall average evidence of positive or purifying selection across the gene region, codon-based Z-tests can be implemented with the Nei-Gojobori method and variance estimated using bootstrapping (Hall, 2017).

We interpret that the failure of the gene sequences to determine the same phylogenetic pattern determined in other studies and in particular those with nuclear ribosomal *ITS* (Chapter 3) is likely due to homoplasy and parallel changes in amino acids. There was some improvement in the phylogenetic analysis in comparison with *ITS* and *trnL-F* loci when only third position and introns were included, which is consistent with the result you would expect if selection was indeed occurring. Hao et al. (2009) also found that their gene trees based on *TS* and *DBAT* were not consistent with their *ITS* tree and that *TS* and *DBAT* were overall under purifying selection and hence found a similar result to ourselves. This is possibly due to Paclitaxel production being a recently evolved pathway

unique to *Taxus* and its relatives. Therefore, there has been little time for divergence. Also, because *Taxus* is producing taxanes to serve a function (protecting the tree from herbivores) too many differences in the gene sequence could affect their function.

4.6. **Conclusions**

To our knowledge our sampling of the *TS* region for *Taxus* and allies is the most comprehensive to date and has added many new sequences and taxa to scientific knowledge. Our new primers have also improved the technical efficiency of sequencing the *TS* region. Sequence alignment of both the nucleotide and the protein translation show that there is considerable variation within the gene which could be of value to drug production. We are not able to link the variation to Paclitaxel production and future studies should look more closely at this task as well as expanding the taxon sampling for the remaining exons that were not sampled thoroughly here.

Chapter 5: Final Discussion

5.1. Genetic variation in *Taxus* and Paclitaxel genes

Taxus species produce the anticancer agent called Paclitaxel but little is known about how much natural genetic variation there is in the genus that can be utilised for drug production. Paclitaxel was originally developed to treat various forms of cancer including ovarian, breast and lung cancer but the uses for Paclitaxel have since expanded into other areas. For example, Paclitaxel coated stents have been used to treat coronary artery disease since the mid 2000s (Stone et al., 2004, Misra and Dake, 2019). However, there are some risks to its use as Katsanos et al. (2018) discovered when they determined that the use of Paclitaxel coated balloons and stents in femoropopliteal artery of the lower limbs causes an increased risk of death. There is therefore a pressing need to develop new safer and more effective variants of the secondary metabolite. Approximately 20-30% of patients treated with Paclitaxel experience clinically relevant sensory peripheral neuropathy (Winer et al., 2004) and some ethnic groups are more prone to this problem than others. Thus, there has been some research into its pharmacoethnicity which is ethnic diversity in the drug response or toxicity. Komatsu et al. (2015) identified genetic variants associated with Asian Paclitaxel induced cytotoxicity. Hasegawa et al. (2011) noticed that toxicity to Paclitaxel was more frequently observed in Asian patients.

This thesis has examined genetic variation in *Taxus* and some of the gene regions in the Paclitaxel biosynthesis pathway. It looked at variation in the *TS* and *DBAT* gene across a wide range of samples and found there to be significant numbers of synonymous and non-synonymous mutations across both genes. This finding shows there is considerable genetic diversity for drug developers to work from. Zhang et al. (2011) observed that overexpression of *DBAT* gene leads to an increased yield of Paclitaxel in *T. chinensis* cells.

We are not able, at this point, to link the genetic variants with differences in Paclitaxel production by different species of *Taxus*. However, it is likely that some variants of Paclitaxel found in differing species of *Taxus*, or populations within a species, will be better suited to treat a particular ethnic group of people over another. A novel approach for further investigation would be to compare whether using natural occurring Paclitaxel from different species of *Taxus* to treat patients of differing ethnic groups would reduce

the effect of toxicity, especially to patients of Asian origin. Our research has shown that there are plenty of non-synonymous mutations in the *TS* gene and some in the *DBAT*. We do not know if these amino acid differences will lead to different Paclitaxel molecules, or alter the quantity of the drug produced, but the variation hints to high variability in the way in which Paclitaxel is produced. Köksal et al. (2010) elucidated the crystal structure of taxadiene synthase which will help to advance future protein engineering efforts. It would be interesting to see how *TS* protein structure changes with the variants we have detected.

We were only able to characterise two of the genes involved in Paclitaxel production. Therefore, further research into the other genes involved in Paclitaxel biosynthesis needs to be done. Paclitaxel content varies from tree to tree, depending on variety, location and time of year samples are taken (Vance et al., 1994, Griffin and Hook, 1996, ElSohly et al., 1997, Hook et al., 1999). It is likely that the genetic variation is a major factor causing a variance in Paclitaxel production. Thus, future research needs to compare genetic variants with Paclitaxel content of samples to see how they are related. All the genes in the biochemical pathway need to be sequenced for a large number of well characterised taxa.

5.2. **Taxonomy and systematics**

To correctly assess and compare variation within these genes across the genus it is essential that plants are correctly identified and the relationships between taxa understood. An accurate phylogenetic reconstruction is required to infer Paclitaxel evolution. This thesis has generated results that show support for the recognition of six genera, *Amentotaxus, Austrotaxus, Cephalotaxus, Pseudotaxus, Taxus* and *Torreya* in Taxaceae. It also supports the recognition of nine species of *Taxus* (*T. baccata, T. brevifolia, T. canadensis, T. cuspidata, T. floridana, T. fauna, T. globosa T. sumatrana* and *T. wallichiana*). Grouping of individuals within species is not a straightforward task and a combination of morphological, anatomical and genetic characters are required. There is a need for more field sampling of *Taxus* across its geographical range because our results indicate that patterns of genetic variation are related to geography. The

results from our TCS network analysis groups *T. baccata, T. canadensis, T. cuspidata* and hybrids together. This is not surprising as these grow in broadly similar biomes and latitudes. *Taxus brevifolia, T. floridana* and *T. globosa* which are all North American group closely together. The Asian yew, *T. wallichiana,* and the Indonesian *T. sumatrana* also group together. As a preliminary study, we also took leaf impressions of a subsample of the *Taxus* samples to help with identification. Rows of stomata and papillae were observed and used in association with taxonomic keys to help with identification.

We found evidence to identify hybrids. Sequence heterogeneity was detected in the uncloned *ITS* PCR products/sequences of *T. x media* and *T. x hunnewelliana.* Polymorphisms were detected at several sites and could be mapped to the corresponding bases in their parental sequences. This supports evidence of their hybrid status that *Taxus x hunnewelliana* is a hybrid of *T. cuspidata* x *T. canadensis* and *T. x media* is a hybrid of *T. baccata* x *T. cuspidata.* It is also essential that detailed voucher and passport data are kept so that the source of sequences can be known with confidence. Many of the samples used were from arboreta where such data had been lost.

5.3. Other future directions for research

One of the major limiting factors in Taxol production is the low yield of Taxol from *Taxus* species. Approximately 10,000kg of bark only yields 1kg of Taxol (Vidensek et al., 1990). A tree of around 100 years old will only yield on average 5.74kg of bark (Nadeem et al., 2002). This has led scientists to look at alternative ways to mass produce Taxol. One method is the use a fermentation process involving Taxol-producing endophytic fungi. In 1993, Andrea Stierle and colleagues reported on the production of Taxol by *Taxomyces andreanae,* which is an endophytic ascomycete fungus associated with *Taxus brevifolia* (Stierle et al., 1993). An endophytic fungus is a fungus which lives within a plant without causing it any harm (Wilson, 1995). These findings lead to the subsequent publication of numerous papers looking at the biosynthesis of Taxol and related taxanes by microorganisms. Around 200 endophytic fungi from over 40 fungal genera have been

reported to produce Taxol (Flores-Bustamante et al., 2010, Hao et al., 2013). It is possible that Taxol producing endophytes might be found in plants that do not make Taxol and Li et al. (1996) suggest that plants other than yew may be a source of Taxol produced by endophytes.

However, as of yet industrial scale production of Taxol by microorganisms has not occurred. One reason seems to be a loss of production after repeated cycles of sub-culturing (Staniek et al., 2009). This has led scientists to question the ability of endophytes to produce Taxol independent of their host (Staniek et al., 2009, Heinig et al., 2013). Stierle et al. (1993) found the amount of Taxol taxanes produced by *Taxomyces andreanae* are low and propose improved culturing techniques and the application of genetic engineering to improve production (Stierle et al., 1993). Heinig et al. (2013) set out to find evidence for independent taxane biosynthesis in endophytes. Heinig et al. (2013) were unable to find evidence for independent taxane biosynthesis in any of the endophytes reported in publications to produce Taxol including *T. andeanae.* They argue that the detection of low levels of taxanes in the fungal isolates can be explained by residual taxenes synthesised by the host tree (Heinig et al., 2013). And yet four years earlier Zhang et al. (2009) found that an endophytic fungus, *Cladosporium cladosporioidies,* isolated from the inner bark of *Taxus media,* is capable of producing taxol (Zhang et al., 2009). They suggest that the Taxol yield of *C. cladosporioidies* will increase due to strain improvement and media optimisation.

Taxol producing fungi are naturally resistant to Taxol, even though Taxol is a fungicide against many fungal phytopathogens (Soliman et al., 2013, Kusari et al., 2014). Soliman et al. (2015) conducted a study which suggested that yew formed symbiotic relationships with Taxol producing endophytes to protect itself from wood decaying fungi. The fungi store Taxol in hydrophobic bodies which travel to pathogen entry points where they release the bodies when threatened by a pathogen hence creating a fungicide barrier for the host yew tree (Soliman et al., 2015).

However, despite these studies on endophytes, it is clear that *Taxus* produces Paclitaxel and does so biochemically via the biosynthetic pathways illustrated in Figure 1.1. Advances in the use of *Taxus* for Paclitaxel production will need to better understand

this pathway and natural genetic variation in the genes of the pathway and how the plant production of Paclitaxel interacts with endophytic fungi. The investigation of interactions between tree and fungus for improved production of Paclitaxel is beyond the scope of this thesis but warrants future investigation. A study comparing the genes involved in Paclitaxel production in *Taxus* with fungal endophytes could have interesting results.

5.4. **Conclusions**

Taxus is clearly a taxonomically challenging group of plants and it is essential that its genetic variation is studied and clarified so that variation in paclitaxel can be understood and manipulated in drug development. This thesis has gone some way to achieving that aim and it has also highlighted the large levels of variation in paclitaxel genes that can be utilised in the future development of this highly valuable secondary metabolite.

References

Allington-Jones L. 2015. The Clacton spear: the last one hundred years. *Archaeological Journal,* **172**: 273-296.

Alloatti G, Penna C, Levi RC, Gallo MP, Appendino G, Fenoglio I. 1996. Effects of yew alkaloids and related compounds on guinea-pig isolated perfused heart and papillary muscle. *Life Sciences,* **58**: 845-854.

Appendino G. 1993. Taxol (paclitaxel): Historical and ecological aspects. *Fitoterapia* **64**: 2-25.

Appendino G. 1995. The phytochemistry of the yew tree. *Natural Product Reports,* **12**: 349-360.

Auders AG, Spicer DP, Society RH. 2012. *Royal Horticultural Society Encyclopedia of Conifers: A Comprehensive Guide to Cultivars and Species*, Royal Horticultural Society. Kingsblue Publishing Limited.

Baldwin BG, Sanderson MJ, Porter JM, Wojciechowski MF, Campbell CS, Donoghue MJ. 1995. The ITS region of nuclear ribosomal DNA - A valuable source of evidence on angiosperm phylogeny. *Annals of the Missouri Botanical Garden,* **82**: 247-277.

Bandelt H-J, Forster P, Röhl A. 1999. Median-joining networks for inferring intraspecific phylogenies. *Molecular Biology and Evolution,* **16**: 37-48.

Bristol Myers Squibb. 2011. Package insert for TAXOL® (paclitaxel) injection.

Carolan JC, Hook IL, Chase MW, Kadereit JW, Hodkinson TR. 2006. Phylogenetics of *Papaver* and related genera based on DNA sequences from ITS nuclear ribosomal DNA and plastid trnL intron and trnL-F intergenic spacers. *Annals of Botany,* **98**: 141-55.

Chaw SM, Long H, Wang BS, Zharkikh A, Li WH. 1993. The phylogenetic position of Taxaceae based on 18S ribosomal - RNA sequences. *Journal of Molecular Evolution,* **37**: 624-630.

Cheng Y, Nicolson RG, Tripp K, Chaw S-M. 2000. Phylogeny of Taxaceae and Cephalotaxaceae genera inferred from chloroplast matK gene and nuclear rDNA ITS region. *Molecular Phylogenetics and Evolution,* **14**: 353-365.

Christenhusz MJM, Reveal JL, Farjon A, Gardner MF, Mill RR, Chase MW. 2011. A new classification and linear sequence of extant gymnosperms. *Phytotaxa,* **19**: 55-70.

Christin P-A, Salamin N, Savolainen V, Duvall MR, Besnard G. 2007. C4 photosynthesis evolved in grasses via parallel adaptive genetic changes. *Current Biology,* **17**: 1241-1247.

Clement M, Posada D, Crandall KA. 2000. TCS: a computer program to estimate gene genealogies. *Molecular Ecology,* **9**: 1657-1659.

Clement M, Snell Q, Walker P, Posada D, Crandall K. 2002. TCS: estimating gene genealogies. *ipdps.* IEEE.

Collins D, Mill RR, Moller M. 2003. Species separation of *Taxus baccata*, *T. canadensis*, and *T. cuspidata* (Taxaceae) and origins of their reputed hybrids inferred from RAPD and cpDNA data. *American Journal of Botany,* **90**: 175-182.

Cope EA. 1998. Taxaceae: The genera and cultivated species. *Botanical Review,* **64**: 291-322.

Cragg GM, Schepartz SA, Suffness M, Grever MR. 1993. The taxol supply crisis. New NCI policies for handling the large-scale production of novel natural product anticancer and anti-HIV agents. *Journal of Natural Products,* **56**: 1657-1668.

Cragg GM, Snader KM. 1991. Taxol: the supply issue. *Cancer Cells,* **3**: 233-5.

Croteau R, Ketchum RB, Long R, Kaspera R, Wildung M. 2006. Taxol Biosynthesis and Molecular Genetics. *Phytochemistry Reviews,* **5**: 75-97.

Dempsey D. 2000. *Fators affecting paclitaxel content in yew,* Ph.D thesis, Trinity College The University of Dublin.

Dempsey D, Hook I. 2000. Yew (*Taxus*) species - chemical and morphological variations. *Pharmaceutical Biology,* **38:** 274-80

Dempsey D, O'Flaherty C, Hook I. 2003. Taxus species - Morphological and chemical variations in needles and seeds. *Proceedings of the fourth International Conifer Conference:Conifers for the Future? Acta Horticulturae,* **615**: 223-231.

Denis JN, Correa A, Greene AE. 1990. An improved synthesis of the Taxol side-chain and of RP-56976 *Journal of Organic Chemistry,* **55**: 1957-1959.

Denis JN, Greene AE, Guenard D, Gueritte-Voegelein F, Mangatal L, Potier P. 1988. Highly efficient, practical approach to natural taxol. *Journal of the American Chemical Society,* **110**: 5917-5919.

Doyle J, Doyle J. 1987. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochemical Bulletins,* **19**: 11-15.

Eckenwalder JE. 2009. *Conifers of the World: The Complete Reference*, Timber Press, Portland, London.

Elder Jr JF, Turner BJ. 1995. Concerted evolution of repetitive DNA sequences in eukaryotes. *The Quarterly Review of Biology,* **70**: 297-320.

Elkassaby YA, Yanchuk AD. 1994. Genetic diversity, differentiation, and genetic diversity, differentiation, and inbreeding in Pacific yew from British Columbia. *Journal of Heredity,* **85**: 112-117.

Elpe C, Knopf P, Stutzel T, Schulz C. 2018. Diversity and evolution of leaf anatomical characters in Taxaceae s.l.-fluorescence microscopy reveals new delimitating characters. *Journal of Plant Research,* **131**: 125-141.

ElSohly HN, Croom EM, Kopycki WJ, Joshi AS, McChesney JD. 1997. Diurnal and seasonal effects on the taxane content of the clippings of certain *Taxus* cultivars. *Phytochemical Analysis,* **8**: 124-129.

Elwes HJ, Henry A. 1906. *The Trees of Great Britain,* Edinburgh, SR Publishers Ltd.

Farjon A. 1998. *World Checklist and Bibliography of Conifers (Second Edition)*, Royal Botanic Gardens, Kew.

Farjon A. 2010a. *A Handbook of the World's Conifers Volume 2*, Brill, Leiden, Boston.

Farjon A. 2010b. *Handbook of the World's Conifers, Vol 1*, Brill, Leiden, Boston.

Farjon A. 2017. *A Handbook of the World's Conifers (2 vols.): Revised and Updated Edition*, Brill, Leiden, Boston.

Feng Y, Oh S-H, Manos PS. 2005. Phylogeny and historical biogeography of the genus Platanus as inferred from nuclear and chloroplast DNA. *Systematic Botany,* **30**: 786-799.

Flores-Bustamante ZR, Rivera-Orduna FN, Martinez-Cardenas A, Flores-Cotera LB. 2010. Microbial paclitaxel: advances and perspectives. *Journal of Antibiotics,* **63**: 460-467.

Florin R. 1951. *Evolution in cordaites and conifers*, Acta Horti Berginal, **17**:259-388.

Fu L, Li N, Mill R. 1999. Taxaceae. *Flora of China,* **4**: 89-96.

Gawel NJ, Jarret RL. 1991. A modifies CTAB DNA extraction procedure for *Musa* and *Ipomoea. Plant Molecular Biology Reporter*: 262-266.

Ghimire B, Heo K. 2014. Cladistic analysis of Taxaceae s. l. *Plant Systematics and Evolution,* **300**: 217-223.

Goodman J, Walsh V. 2001. *The Story of Taxol*. 2001 ed., Cambridge University Press.

Graf E, Boeddeker H. 1956. Zur kenntnis der B-dimethylamino-hydrozimtsaure. *Archiv der Pharm und Berichte der Deutsch Pharmaz Ges*: 364-370.

Griffin J, Hook I. 1996. Taxol content of Irish Yews. *Planta Medica,* **62**: 370-372.

Guo BH, Kai GY, Jin HB, Tang KX. 2006. Taxol synthesis. *African Journal of Biotechnology,* **5**: 015-020.

Hageneder F. 2005. *The Living wisdom of trees,* London, Duncan Baird Publishers.

Hageneder F. 2007. *Yew - a history,* Gloucestershire, Sutton Publishing.

Hall BG. 2017. *Phylogenetic Trees Made Easy. A How-To Manual* Oxford University Press, Sunderland, Massachusetts.

Hall TA. 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic acids symposium series*, **41**: 95-98.

Hao C, Yang L, Huang B. 2009. Molecular evolution of paclitaxel biosynthetic genes TS and DBAT of *Taxus* species. *Genetica,* **135**: 123-35.

Hao DC, Huang B, Yang L. 2008a. Phylogenetic relationships of the genus *Taxus* inferred from chloroplast intergenic spacer and nuclear coding DNA. *Biological & Pharmaceutical Bulletin,* **31**: 260-265.

Hao DC, Xiao PG, Huang BL, Ge GB, Yang L. 2008b. Interspecific relationships and origins of Taxaceae and Cephalotaxaceae revealed by partitioned Bayesian analyses of chloroplast and nuclear DNA sequences. *Plant Systematics and Evolution,* **276**: 89-104.

Hao X, Pan J, Zhu X. 2013. Taxol producing fungi. In: Ramawat KG, Mérillon J-M eds. *Natural Products: Phytochemistry, Botany and Metabolism of Alkaloids, Phenolics and Terpenes.* Berlin, Heidelberg, Springer Berlin Heidelberg.

Hartzell HJ. 1991. *The yew tree:a thousand whispers,* Oregon, Hulogosi.

Hasegawa Y, Kawaguchi T, Kubo A, Ando M, Shiraishi J, Isa S, Tsuji T, Tsujino K, Ou SH, Nakagawa K, Takada M. 2011. Ethnic difference in hematological toxicity in patients with non-small cell lung cancer treated with chemotherapy: a pooled analysis on Asian versus non-Asian in phase II and III clinical trials. *Journal of Thoracic Oncology,* **6**: 1881-8.

Hefner, J., Ketchum, R.E. and Croteau, R., 1998. Cloning and functional expression of a cDNA encoding geranylgeranyl diphosphate synthase from *Taxus canadensis* and assessment of the role of this prenyltransferase in cells induced for taxol production. *Archives of biochemistry and biophysics*, **360**:62-74.

Heinig U, Scholz S, Jennewein S. 2013. Getting to the bottom of Taxol biosynthesis by fungi. *Fungal Diversity,* **60**: 161-170.

Hodkinson TR, Chase MW, Takahashi C, Leitch IJ, Bennett MD, Renvoize SA. 2002. The use of DNA sequencing (ITS and trnL-F), AFLP, and fluorescent *in situ* hybridization to study allopolyploid *Miscanthus* (Poaceae). *American Journal of Botany,* **89**: 279-286.

Hodkinson TR, Waldren S, Parnell JAN, Kelleher CT, Salamin K, Salamin N. 2007. DNA banking for plant breeding, biotechnology and biodiversity evaluation. *Journal of Plant Research,* **120**: 17-29.

Holmes FA, Walters RS, Theriault RL, Buzdar AU, Frye DK, Hortobagyi GN, Forman AD, Newton LK, Raber MN. 1991. Phase II trial of Taxol, an active drug in the treatment of metastatic breast cancer. *JNCI: Journal of the National Cancer Institute,* **83**: 1797-1805.

Hook I, Poupat C, Ahond A, Guenard D, Gueritte F, Adeline MT, Wang XP, Dempsey D, Breuillet S, Potier P. 1999. Seasonal variation of neutral and basic taxoid contents in shoots of European Yew (*Taxus baccata*). *Phytochemistry,* **52**: 1041-1045.

Hultén E, Fries M. 1986. *Atlas of North European vascular plants north of the Tropic of Cancer*, Koeltz Scientific.

Irish Medicines Board. 2007. Medicinal products(control of placing on the market) regulations 2007 (S.I. No. 540 of 2007) Case No:2029946.

Itokawa H, Lee KH. 2003. *Taxus,* London, Taylor and Francis.

Jennewein S, Wildung MR, Chau M, Walker K, Croteau R. 2004. Random sequencing of an induced *Taxus* cell cDNA library for identification of clones involved in Taxol biosynthesis. *Proceedings of the National Academy of Sciences,* **101**: 9149-9154.

Judd WS, Campbell CS, Kellogg EA, Stevens PF, Donoghue MJ. 2016. *Plant Systematics: A phylogenetic Approach*, Sinaver Associates, Sunderland Massachusetts.

Katsanos K, Spiliopoulos S, Kitrou P, Krokidis M, Karnabatidis D. 2018. Risk of death following application of paclitaxel-coated balloons and stents in the femoropopliteal artery of the leg: A systematic review and meta-analysis of randomized controlled trials. *Journal of the American Heart Association,* **7**: e011245.

Kelsey RG, Vance NC. 1992. Taxol and cephalomannine concentrations in the foliage and bark of shade grown an sun exposed *Taxus brevifolia* trees. *Journal of Natural Products,* **55**: 912-917.

Kilmartin L. 2002. The Molecular Systematics of *Taxus* and Related Genera M.Sc thesis, Trinity College, The University of Dublin.

Koepp AE, Hezari M, Zajicek J, Vogel BS, LaFever RE, Lewis NG, Croteau R. 1995. Cyclization of Geranylgeranyl Diphosphate to Taxa-4 (), 11 () -diene is the committed step of Taxol biosynthesis in Pacific Yew. *Journal of Biological Chemistry,* **270**: 8686-8690.

Köksal M, Jin Y, Coates RM, Croteau R, Christianson DW. 2010. Taxadiene synthase structure and evolution of modular architecture in terpene biosynthesis. *Nature,* **469**: 116.

Komatsu M, Wheeler HE, Chung S, Low S-K, Wing C, Delaney SM, Gorsic LK, Takahashi A, Kubo M, Kroetz DL. 2015. Pharmacoethnicity in paclitaxel-induced sensory peripheral neuropathy. *Clinical Cancer Research,* **21**: 4337-4346.

Korneliussen TS, Moltke I, Albrechtsen A, Nielsen R. 2013. Calculation of Tajima's D and other neutrality test statistics from low depth next-generation sequencing data. *BMC Bioinformatics,* **14**: 289.

Kumar S, Stecher G, Tamura K. 2016. MEGA 7: Molecular Evolutionary Genetics Analysis version 7 for bigger data sets. *Molecular Biology and Evolution,* **33**: 1870-1874.

Kusari S, Singh S, Jayabaskaran C. 2014. Rethinking production of Taxol® (paclitaxel) using endophyte biotechnology. *Trends in Biotechnology,* **32**: 304-311.

Kwit C, Horvitz CC, Platt WJ. 2004. Conserving slow-growing, long-lived tree species: input from the demography of a rare understory conifer, *Taxus floridana*. *Conservation Biology,* **18**: 432-443.

Lanker U, Malik AR, Gupta NK, Butola JS. 2010. Natural regeneration status of the endangered medicinal plant, *Taxus baccata* Hook. F. syn. *T. wallichiana*, in northwest Himalaya. *International Journal of Biodiversity Science, Ecosystem Services & Management,* **6**: 20-27.

Lavabre JE, García García D. 2015. Geographic consistency in the seed dispersal patterns of *Taxus baccata* L. in the Iberian Peninsula. *Forest Systems*, **24:** e040.

Leigh JW, Bryant D. 2015. Popart: full-feature software for haplotype nextwork construction. *Methods Ecology Evolution,* **6**: 1110-1116.

Li J, Davis CC, Del Tredici P, Donoghue MJ. 2001. Phylogeny and biogeography of *Taxus* (Taxaceae) inferred from sequences of the internal transcribed spacer region of nuclear ribosomal DNA *Harvard Papers in Botany,* **6**: 267-274.

Li JY, Strobel G, Sidhu R, Hess WM, Ford EJ. 1996. Endophytic taxol-producing fungi from bald cypress, *Taxodium distichum*. *Microbiology-Uk,* **142**: 2223-2226.

Li N, Fu LK. 1997. Notes on gymnosperms I. Taxonomic treatments of some Chinese conifers. *Novon,* **7**: 261-264.

Liston A, Robinson WA, Oliphant JM, Alvarez-Buylla ER. 1996. Length variation in the nuclear ribosomal DNA internal transcribed spacer region of non-flowering seed plants. *Systematic Botany,* **21**: 109-120.

Lucas H. 1856. Ueber ein in den Blättern von *Taxus baccata* L. enthaltenes alkaloid (das Taxin). *Archiv der Pharmazie,* **135**: 145-149.

MacCoitir N. 2003. *Irish Trees myths, Legend & folklore,* Cork, The Collins Press.

Mattina M, MacEachern G. 1994. Extraction, purification by solid-phase extraction and high-performance liquid chromatographic analysis of taxanes from ornamental *Taxus* needles. *Journal of Chromatography A,* **679**: 269-275.

McGuire WP, Rowinsky EK, Rosenshein NB, et al. 1989. Taxol: A unique antineoplastic agent with significant activity in advanced ovarian epithelial neoplasms. *Annals of Internal Medicine,* **111**: 273-279.

Misra S, Dake MD. 2019. Paclitaxel-based therapies for patients with peripheral artery disease. *Circulation,* **139**: 1565-1567.

Mohapatra K, Sehgal R, Sharma R, Mohapatra T. 2009. Genetic analysis and conservation of endangered medicinal tree species *Taxus wallichiana* in the Himalayan region. *New Forests,* **37**: 109-121.

Mullis KB, Faloona FA. 1987. Specific synthesis of DNA invitro via a polymerase-catalyzed chain-reaction. *Methods in Enzymology,* **155**: 335-350.

Nadeem M, Rikhari HC, Kumar A, Palni LMS, Nandi SK. 2002. Taxol content in the bark of Himalayan Yew in relation to tree age and sex. *Phytochemistry,* **60**: 627-631.

Nei M, Gojobori T. 1986. Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. *Molecular Biology and Evolution,* **3**: 418-426.

Nei M, Kumar S. 2000. *Molecular Evolution and Phylogenetics*, Oxford University Press, Oxford, UK.

Nelson, E.C & Walsh, W.F 1993. *Trees of Ireland: native and naturalized.* Lilliput Press, Dublin.

Nims E, Dubois CP, Roberts SC, Walker EL. 2006. Expression profiling of genes involved in paclitaxel biosynthesis for targeted metabolic engineering. *Metabolic Engineering,* **8**: 385-394.

Parnell J, Curtis T. 2012. *Webb's An Irish Flora*, Cork University Press.

Perrin PM. 2002. *The Ecology of Yew (Taxus baccata) in Ireland,* Ph.D thesis, Trinity College The University of Dublin.

Pilger R. 1903. *Taxaceae* Leipzig Berlin W. Engelmann

Pilger R. 1926. Taxodiaceae. *Die Naturlichen Pflanzenfamilien,* **13**.

Poupat C, Hook I, Gueritte F, Ahond A, Guenard D, Adeline MT, Wang XP, Dempsey D, Breuillet S, Potier P. 2000. Neutral and basic taxoid contents in the needles of *Taxus* species. *Planta Medica,* **66**: 580-584.

Price RA. 2003. Generic and familial relationships of the Taxaceae from rbcL and matK sequence comparisons. In: Mill RR ed. *Proceedings of the Fourth International Conifer Conference: Conifers for the Future?* Acta Horticulturae ***615***: 235-237.

Rehder A. 1923. *Taxus media* hybrid. *Journal of the Arnold Arboretum,* **4**: 106-109.

Rehder A. 1925. *Taxus hunnewelliana* hybrid *Journal of Arnold Arboretum,* **6**: 201-202.

Ronquist F, Huelsenback JP. 2003. MRBAYES 3: Bayesian phylogenetic interence under mixed models. *Bioinformatics,* **19**: 1572-1574.

Rowinsky EK, Cazenave LA, Donehower RC. 1990. Taxol: a novel investigational antimicrotubule agent. *JNCI: Journal of the National Cancer Institute,* **82**: 1247-1259.

Sarvella P, Meyer JR, Owings A. 1961. A scotch tape method for counting or measuring stomata. *Crop Science,* **1**: 81-82.

Shah A, Li D-Z, Gao L-M, Li H-T, Möller M. 2008a. Genetic diversity within and among populations of the endangered species *Taxus fuana* (Taxaceae) from Pakistan and implications for its conservation. *Biochemical Systematics and Ecology,* **36**: 183-193.

Shah A, Li D-Z, Möller M, Gao L-M, Hollingsworth ML, Gibby M. 2008b. Delimitation of *Taxus fuana* Nan Li & RR Mill (Taxaceae) based on morphological and molecular data. *Taxon,* **57**: 211-222.

Snow B, Snow D. 2010. *Birds and berries*, A&C Black, London.

Soliman Sameh SM, Greenwood John S, Bombarely A, Mueller Lukas A, Tsao R, Mosser Dick D, Raizada Manish N. 2015. An endophyte constructs fungicide-containing extracellular barriers for its host plant. *Current Biology,* **25**: 2570-2576.

Soliman SSM, Trobacher CP, Tsao R, Greenwood JS, Raizada MN. 2013. A fungal endophyte induces transcription of genes encoding a redundant fungicide pathway in its host plant. *Bmc Plant Biology,* **13**.

Spector T, Thomas P, Determann R. 2011 *Taxus floridana.The IUCN Red List of ThreatenedSpecies:* http://dx.doi.org/10.2305/IUCN.UK.20112.RLTS.T30965A9584928.en (accessed on 1st November 2018).

Spjut RW. 2007a. A phytogeographical analysis of *Taxus* (Taxaceae) based on leaf anatomical characters. *Journal of the Botanical Research Institute of Texas,* **1**: 291-332.

Spjut RW. 2007b. Taxonomy and nomenclature of *Taxus* (Taxaceae). *Journal of the Botanical Research Institute of Texas,* **1**: 203-289.

Staniek A, Woerdenbag HJ, Kayser O. 2009. *Taxomyces andreanae*: a presumed paclitaxel producer demystified? *Planta Medica,* **75**: 1561-1566.

Stierle A, Strobel G, Donald S. 1993. Taxol and Taxane production by *Taxomyces andreanae,* an Endophytic Fungus of Pacific Yew. *Science* **260**: 214-216.

Stone GW, Ellis SG, Cox DA, Hermiller J, O'Shaughnessy C, Mann JT, Turco M, Caputo R, Bergin P, Greenberg J. 2004. One-year clinical results with the slow-release, polymer-based, paclitaxel-eluting TAXUS stent: the TAXUS-IV trial. *Circulation,* **109**: 1942-1947.

Sun Y, Skinner DZ, Liang GH, Hulbert SH. 1994. Phylogenetic analysis of sorghum and related taxa using internal transcribed spacers of nuclear ribosomal DNA. *Theoretical and Applied Genetics,* **89**: 26-32.

Taberlet P, Gielly L, Pautou G, Bouve J. 1991. Universal primers for amplification of three non-coding regions of chloroplast DNA. *Plant Molecular Biology* **17:** 1105-1109.

The Plant List 2013. Version1.1 Published on the Internet; http://www.theplantlist.org/ (accessed 3rd September 2018).

Thomas P. 2011. *Taxus contorta*. *The IUCN Red List of Threatened Species*: http://dx.doi.org/10.2305/IUCN.UK.2011-2.RLTS.T39147A10170545.en (accessed on 1st November 2018).

Thomas P. 2013a. *Taxus brevifolia*. *The IUCN Red List of Threatened Species*: http://dx.doi.org/10.2305/IUCN.UK.2013-1.RLTS.T34041A2841142.en (accessed on 1st November 2018).

Thomas P. 2013b. *Taxus globosa*. *The IUCN Red List of Threatened Species*: http://dx.doi.org/10.2305/IUCN.UK.2013-1.RLTS.T30724A2795235.en (accessed on 1st November 2018).

Thomas P, Farjon A. 2011. *Taxus wallichiana*. *The IUCN Red List of Threatened Species*: http://dx.doi.org/10.2305/IUCN.UK.2011-2.RLTS.T46171879A9730085.en (accessed on 1st November 2018)

Thomas P, Li N, Christian T. 2013. *Taxus chinensis*. *The IUCN Red List of Threatened Species*:http://dx.doi.org/10.2305/IUCN.UK.2013-1.RLTS.T42548A2987120.en (accessed on 1st November 2018).

Thomas P, Polwart A. 2003. *Taxus baccata* L. *Journal of Ecology,* **91**: 489-524.

Toyomasu T, Sassa T. 2010. *Diterpenes. Comprehensive Natural Products II, Elsevier Science* **1**: 643-672.

Trapp SC, Croteau RB. 2001. Genomic organization of plant terpene synthases and molecular evolutionary implications. *Genetics,* **158**: 811-832.

Tyler VE. 1960. Note on the occurrence of taxine in *Taxus brevifolia*. *Journal of the American Pharmaceutical Association,* **49**: 683-684.

van Rozendaal ELM, Lelyveld GP, van Beek TA. 2000. Screening of the needles of different yew species and cultivars for paclitaxel and related taxoids. *Phytochemistry,* **53**: 383-389.

Vance NC, Kelsey RG, Sabin TE. 1994. Seasonal and tissue variation in taxane concentrations of Taxus brevifolia. *Phytochemistry,* **36**: 1241-1244.

Vidensek N, Lim P, Campbell A, Carlson C. 1990. Taxol content in bark, wood, root, leaf, twig, and seedling from several Taxus species. *Journal of Natural Products,* **53**: 1609-1610.

Walker K, Croteau R. 2000. Molecular cloning of a 10-deacetylbaccatin III-10-O-acetyl transferase cDNA from Taxus and functional expression in Escherichia coli. *Proceedings of the National Academy of Sciences of the United States of America,* **97**: 583-587.

Walker K, Croteau R. 2001. Taxol biosynthetic genes. *Phytochemistry,* **58**: 1-7.

Wani MC, Taylor HL, Wall ME, Coggon P, McPhail AT. 1971. Plant antitumor agents .6. isolation and structure of Taxol, a novel antileukemic and antitumor agent frm T.brevifolia. *Journal of the American Chemical Society,* **93**: 2325-&.

Welch H, Gordon H, pool Wcd. 1993. *The World Checklist of Conifers,* Buchenhill, Bromyard, Herefordshire, Landsman Bookshop on behalf of the World Conifer Datapool.

Wendel JF, Doyle JJ. 1998. Phylogenetic incongruence: window into genome history and molecular evolution. *Molecular Systematics of Plants II,* 265-296. Springer, Boston, Massachusetts.

Wendel JF, Schnabel A, Seelanan T. 1995. An unusal ribosomal DNA-sequence from *Gossypium gossypioides* reveals ancient,cryptic,intergenomic introgression. *Molecular Phylogenetics and Evolution,* **4**: 298-313.

Wildung, M.R. and Croteau, R., 1996. A cDNA clone for taxadiene synthase, the diterpene cyclase that catalyzes the committed step of taxol biosynthesis. *Journal of Biological Chemistry*, **271**: 9201-9204.

Williamson R. 1978. *The Great Yew Forest: the Natural History of Kingley Vale*, MacMillan, London.

Wilson CR, Sauer JM, Hooser SB. 2001. Taxines: a review of the mechanism and toxicity of yew (Taxus spp.) alkaloids. *Toxicon,* **39**: 175-185.

Wilson D. 1995. Endophyte: the evolution of a term, and clarification of its use and definition. *Oikos,* **73**: 274-276.

Winer EP, Berry DA, Woolf S, Duggan D, Kornblith A, Harris LN, Michaelson RA, Kirshner JA, Fleming GF, Perry MC. 2004. Failure of higher-dose paclitaxel to improve outcome in patients with metastatic breast cancer: cancer and leukemia group B trial 9342. *Journal of Clinical Oncology,* **22**: 2061-2068.

Witherup KM, Look SA, Stasko MW, Ghiorzi TJ, Muschik GM, Cragg GM. 1990. *Taxus* spp. needles contain amounts of taxol comparable to the bark of *Taxus brevifolia*: analysis and isolation. *Journal of Natural Products,* **53**: 1249-1255.

Yang L, Christian T, Li N. 2013. *Taxus mairei*. *The IUCN Red List of Threatened Species*: http://dx.doi.org/10.2305/IUCN.UK.2013-1.RLTS.T191659A1991533.en (accessed on 1st November 2018).

Zhang J-T, Ru W. 2010. Population characteristics of endangered species *Taxus chinensis* var. mairei and its conservation strategy in Shanxi, China. *Population ecology,* **52**: 407-416.

Zhang P, Li S-T, Liu T-T, Fu C-H, Zhou P-P, Zhao C-F, Yu L-J. 2011. Overexpression of a 10-deacetylbaccatin III-10 β-O-acetyltransferase gene leads to increased taxol yield in cells of *Taxus chinensis*. *Plant Cell, Tissue and Organ Culture (PCTOC),* **106**: 63-70.

Zhang P, Zhou PP, Yu LJ. 2009. An endophytic taxol-producing fungus from *Taxus media*, *Cladosporium cladosporioides* MD2. *Current Microbiology,* **59**: 227-232.

Supplementary figures and tables

Supplementary table 3.1. Maximum Likelihood fits of 24 different nucleotide substitution models for the ITS dataset.

Model	#Param	BIC	AICc	lnL
T92+G	272	7826.990346	5140.50711	-2297.738457
HKY+G	274	7827.654069	5121.424897	-2286.189755
T92+I	272	7836.023465	5149.540229	-2302.255017
HKY+I	274	7836.375286	5130.146114	-2290.550363
T92+G+I	273	7837.947438	5141.59122	-2297.276721
HKY+G+I	275	7838.344952	5122.242854	-2285.594914
TN93+G	275	7839.000949	5122.89885	-2285.922912
HKY	273	7839.858514	5143.502296	-2298.232259
T92	271	7841.973816	5165.36359	-2311.170474
TN93+I	275	7847.847438	5131.74534	-2290.346157
TN93+G+I	276	7849.97207	5123.997074	-2285.468191
TN93	274	7853.762101	5147.532929	-2299.243771
K2+G	271	7869.388791	5192.778565	-2324.877962
GTR+G	278	7873.236659	5127.515949	-2285.219921
K2+I	271	7875.318038	5198.707811	-2327.842585
K2+G+I	272	7876.40138	5189.918144	-2322.443974
K2	270	7881.568055	5214.830867	-2336.907876
GTR+I	278	7881.66257	5135.94186	-2289.432877
GTR+G+I	279	7884.220201	5128.626677	-2284.77141
GTR	277	7887.974236	5152.126369	-2298.528992
JC+G	270	7918.864363	5252.127175	-2355.55603
JC+I	270	7924.687587	5257.950398	-2358.467642
JC+G+I	271	7925.948801	5249.338575	-2353.157967
JC	269	7928.470271	5271.606148	-2366.299266

Note: Models with the lowest BIC scores (Bayesian Information Criterion) are considered to describe the substitution pattern the best. For each model, AICc value (Akaike Information Criterion, corrected), Maximum Likelihood value (lnL), and the number of parameters (including branch lengths) are also presented. Non-uniformity of evolutionary rates among sites may be modeled by using a discrete Gamma distribution (+G) with 5 rate categories and by assuming that a certain fraction of sites are evolutionarily invariable (+I). Abbreviations: GTR: General Time Reversible; HKY: Hasegawa-Kishino-Yano; TN93: Tamura-Nei; T92: Tamura 3-parameter; K2: Kimura 2-parameter; JC: Jukes-Cantor.

Supplementary table 3.2. Maximum Likelihood fits of 24 different nucleotide substitution models for the trnL-F dataset

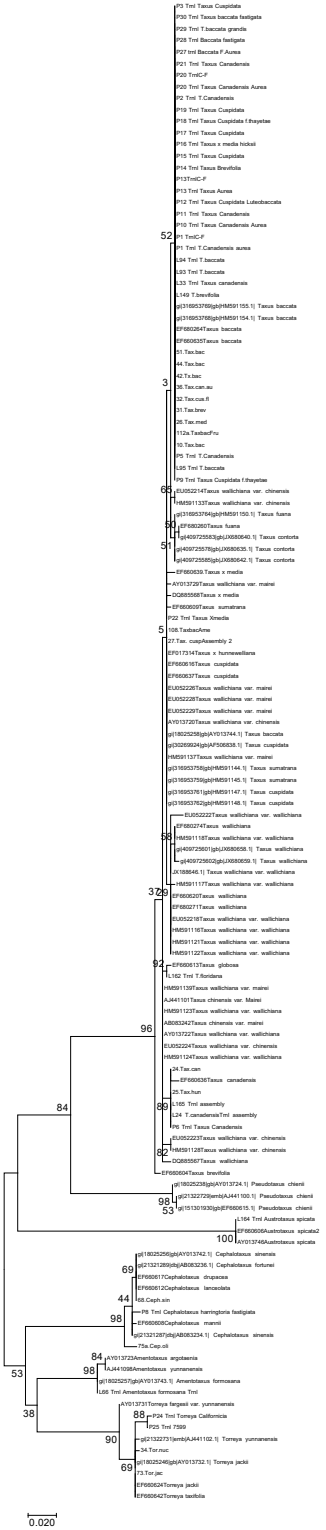
Model	#Param	BIC	AICc	lnL
T92+G	192	3727.006771	2109.968733	-861.8875809
T92	191	3728.127877	2119.500542	-867.664883
T92+I	192	3730.186651	2113.148613	-863.4775208
T92+G+I	193	3737.44027	2111.991649	-861.8875811
HKY+G	194	3746.746351	2112.887267	-861.3238729
HKY	193	3747.80431	2122.355689	-867.0696011
JC+G	190	3748.943337	2148.726824	-883.289362
HKY+I	194	3749.882717	2116.023633	-862.8920558
JC+I	190	3752.439347	2152.222833	-885.0373665
JC	189	3753.586917	2161.781346	-890.8279009
TN93+G	195	3756.147777	2113.87835	-860.8078369
TN93	194	3756.70311	2122.844026	-866.3022522
HKY+G+I	195	3757.181862	2114.912435	-861.3248793
K2+G	191	3757.917928	2149.290592	-882.5599083
TN93+I	195	3758.864046	2116.594619	-862.1659715
JC+G+I	191	3759.278152	2150.650816	-883.2400201
K2	190	3759.466443	2159.24993	-888.5509149
K2+I	191	3761.450805	2152.823469	-884.3263468
GTR+I	198	3763.487881	2095.988144	-848.8276421
GTR	197	3765.215826	2106.126072	-854.9083633
GTR+G	198	3765.988505	2098.488767	-850.0779537
TN93+G+I	196	3766.243339	2115.563689	-860.6388691
K2+G+I	192	3768.337699	2151.299661	-882.5530447
GTR+G+I	199	3774.66318	2098.753578	-849.1985424

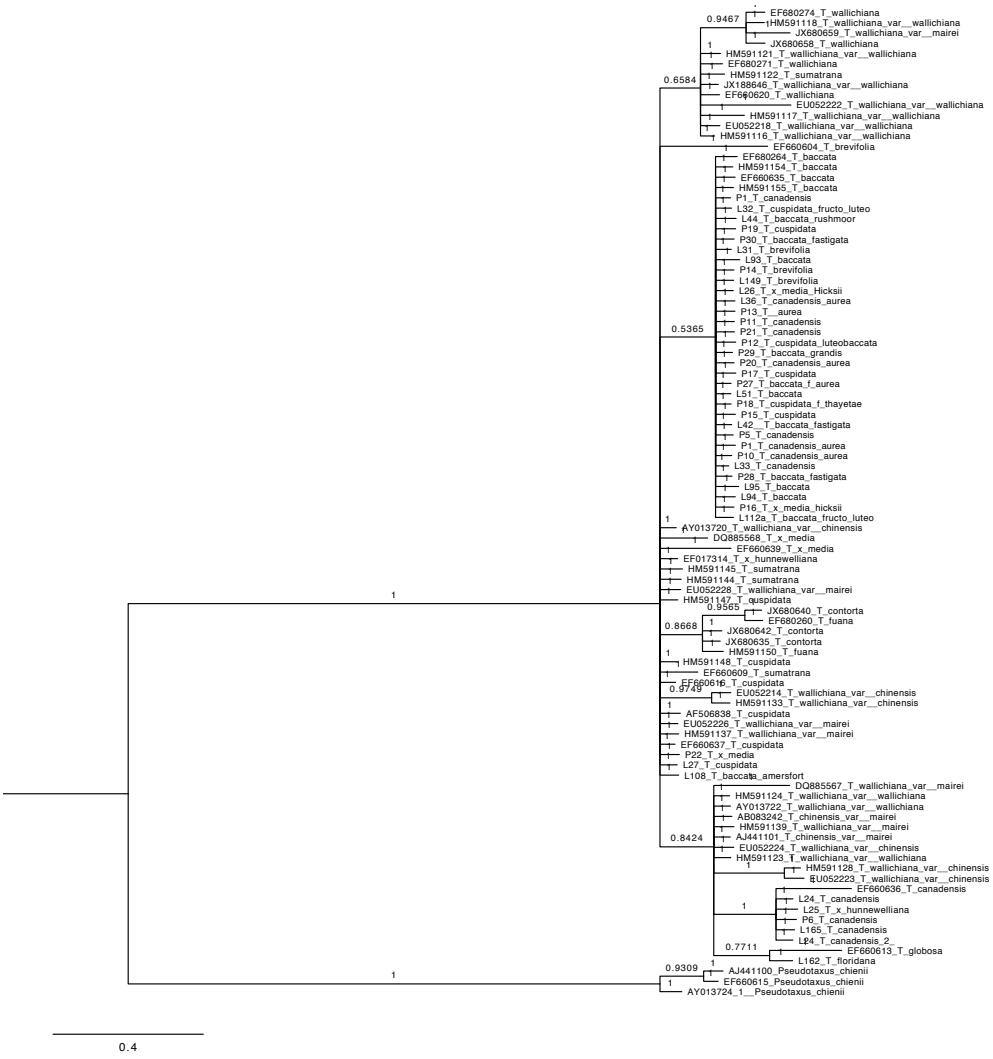
Note: Models with the lowest BIC scores (Bayesian Information Criterion) are considered to describe the substitution pattern the best. For each model, AICc value (Akaike Information Criterion, corrected), Maximum Likelihood value (lnL), and the number of parameters (including branch lengths) are also presented. Non-uniformity of evolutionary rates among sites may be modeled by using a discrete Gamma distribution (+G) with 5 rate categories and by assuming that a certain fraction of sites are evolutionarily invariable (+I). Abbreviations: GTR: General Time Reversible; HKY: Hasegawa-Kishino-Yano; TN93: Tamura-Nei; T92: Tamura 3-parameter; K2: Kimura 2-parameter; JC: Jukes-Cantor.

Supplementary table 3.3. Leaf anatomy data for Taxus species.

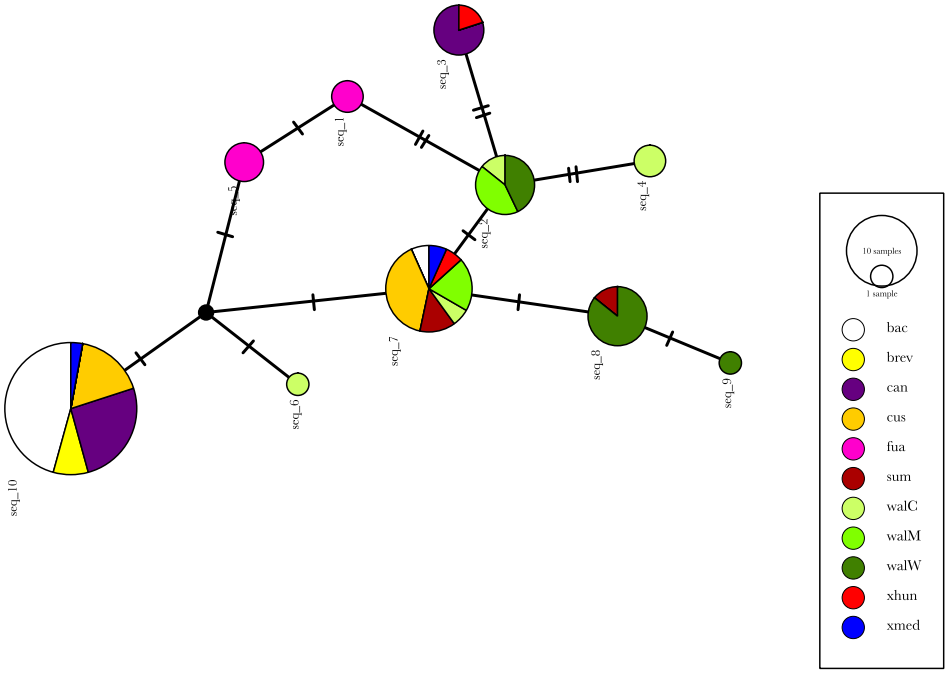
Sample no.	Sample name	Rows of stomata	Midrib
P9	*T. cuspidata f.thayerae*	9	Papillae absent
P11	*T. canadensis*	5 to 6	Papillae present at tip
P12	*T. cuspidata "Luteobaccata"*	6	Papillae present near base of leaf only
P13	*T. canadensis "Aurea"*	4	Papillae present near tip of leaf only
P14	*T. brevifolia*	5 to 7	Papillae present near tip of leaf only
P15	*T. cuspidata*	8 to 9	Papillae present. Papillae absent at base of leaf
P16	*T. xmedia "Hicksii"*	7	Papillae absent
P17	*T. cuspidata*	6	Papillae present at tip of leaf only
P19	*T. cuspidata*	6	Papillae present near base of leaf only
P29	*T. baccata "Grandis"*	7	Papillae present near tip of leaf only
T1	*T. baccata*	8	Papillae present
St. Doolaghs	*T. baccata*	7 to 8	Papillae absent

*Supplementary Figure 3.1. Maximum likelihood tree for a) trnL-F (including Taxus, Pseudotaxus, Austrotaxus, Amentotaxus, Cephalotaxus and Torreya; and b) trnL-F (including Taxus and Pseudotaxus only). Bootstrap values shown above the branches.a)_*



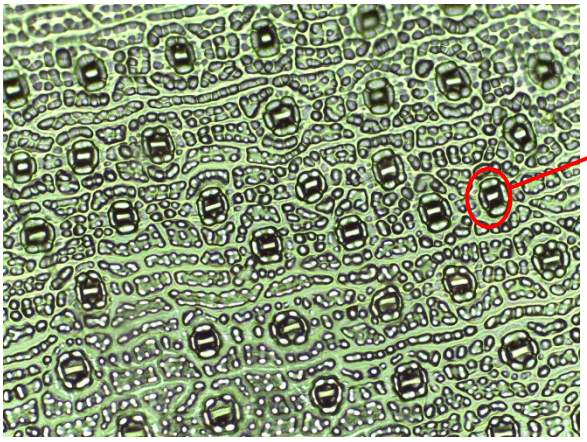

EF680274_T_wallichiana
HM591118_T_wallichiana_var__wallichiana
JX680659_T_wallichiana_var__mairei
JX680658_T_wallichiana
HM591121_T_wallichiana_var__wallichiana
EF680271_T_wallichiana
HM591122_T_sumatrana
JX188646_T_wallichiana_var__wallichiana
EF660620_T_wallichiana
EU052222_T_wallichiana_var__wallichiana
HM591117_T_wallichiana_var__wallichiana
EU052218_T_wallichiana_var__wallichiana
HM591116_T_wallichiana_var__wallichiana
EF660604_T_brevifolia
EF680264_T_baccata
HM591154_T_baccata
EF660635_T_baccata
HM591155_T_baccata
P1_T_canadensis
L32_T_cuspidata_fructo_luteo
L44_T_baccata_rushmoor
P19_T_cuspidata
P30_T_baccata_fastigata
L31_T_brevifolia
L93_T_baccata
P14_T_brevifolia
L149_T_brevifolia
L26_T_x_media_Hicksii
L36_T_canadensis_aurea
P13_T__aurea
P11_T_canadensis
P21_T_canadensis
P12_T_cuspidata_luteobaccata
P29_T_baccata_grandis
P20_T_canadensis_aurea
P17_T_cuspidata
P27_T_baccata_f_aurea
L51_T_baccata
P18_T_cuspidata_f_thayetae
P15_T_cuspidata
L42__T_baccata_fastigata
P5_T_canadensis
P1_T_canadensis_aurea
P10_T_canadensis_aurea
L33_T_canadensis
P28_T_baccata_fastigata
L95_T_baccata
L94_T_baccata
P16_T_x_media_hicksii
L112a_T_baccata_fructo_luteo
AY013720_T_wallichiana_var__chinensis
DQ885568_T_x_media
EF660639_T_x_media
EF017314_T_x_hunnewelliana
HM591145_T_sumatrana
HM591144_T_sumatrana
EU052228_T_wallichiana_var__mairei
HM591147_T_cuspidata
JX680640_T_contorta
EF680260_T_fuana
JX680642_T_contorta
JX680635_T_contorta
HM591150_T_fuana
HM591148_T_cuspidata
EF660609_T_sumatrana
EF660616_T_cuspidata
EU052214_T_wallichiana_var__chinensis
HM591133_T_wallichiana_var__chinensis
AF506838_T_cuspidata
EU052226_T_wallichiana_var__mairei
HM591137_T_wallichiana_var__mairei
EF660637_T_cuspidata
P22_T_x_media
L27_T_cuspidata
L108_T_baccata_amersfort
DQ885567_T_wallichiana_var__mairei
HM591124_T_wallichiana_var__wallichiana
AY013722_T_wallichiana_var__wallichiana
AB083242_T_chinensis_var__mairei
HM591139_T_wallichiana_var__mairei
AJ441101_T_chinensis_var__mairei
EU052224_T_wallichiana_var__chinensis
HM591123_T_wallichiana_var__wallichiana
HM591128_T_wallichiana_var__chinensis
EU052223_T_wallichiana_var__chinensis
EF660636_T_canadensis
L24_T_canadensis
L25_T_x_hunnewelliana
P6_T_canadensis
L165_T_canadensis
L24_T_canadensis_2_
EF660613_T_globosa
L162_T_floridana
AJ441100_Pseudotaxus_chienii
EF660615_Pseudotaxus_chienii
AY013724_1__Pseudotaxus_chienii
0.9467
0.6584
0.5365
0.9565
0.8668
0.9749
0.8424
0.7711
0.9309
1
0.4

b

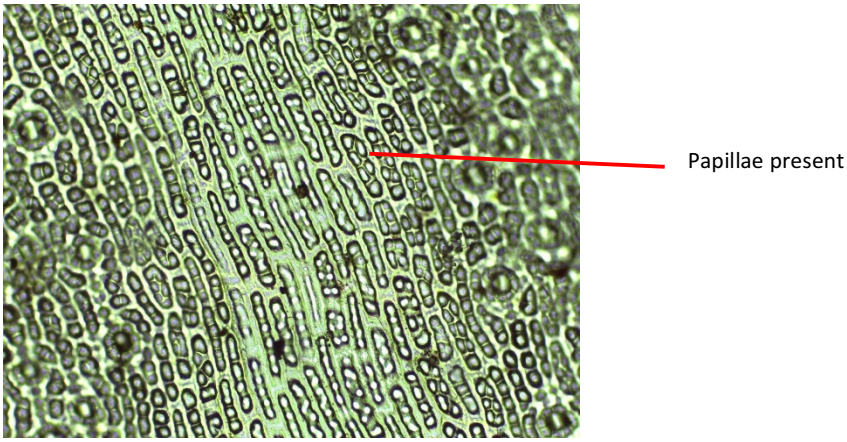


Supplementary Figure 3.2. TrnL intron and trnL-F haplotype TCS analysis

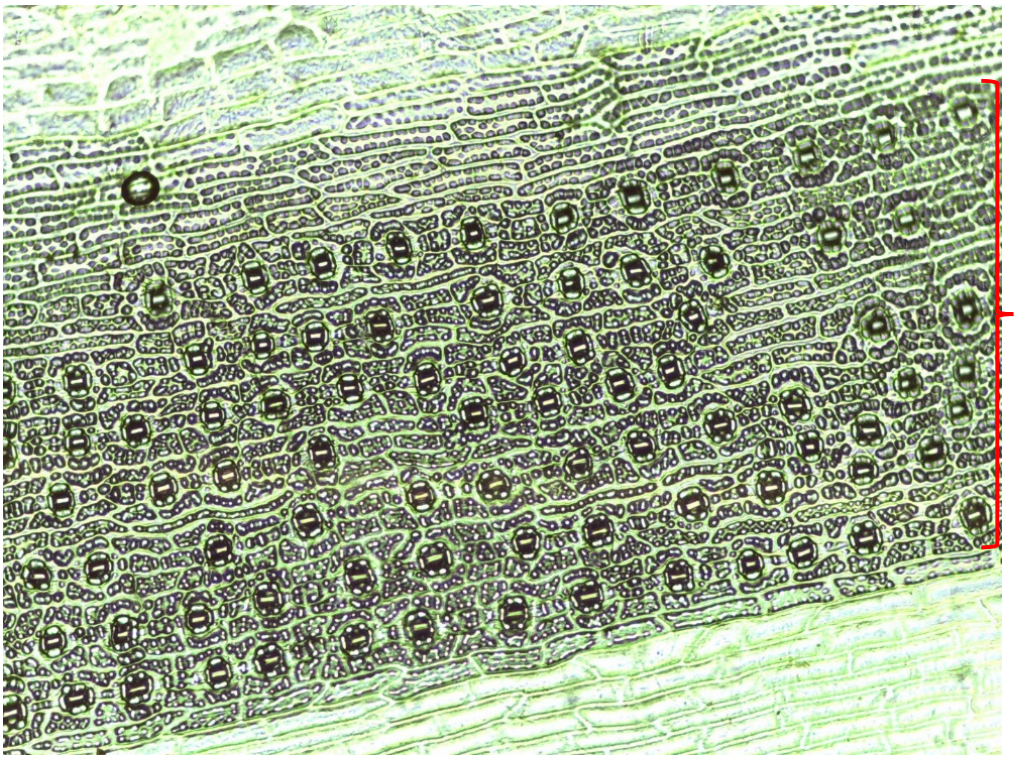
Supplementary Figures 3.3 to 3.16. Leaf anatomy of selected Taxus species.



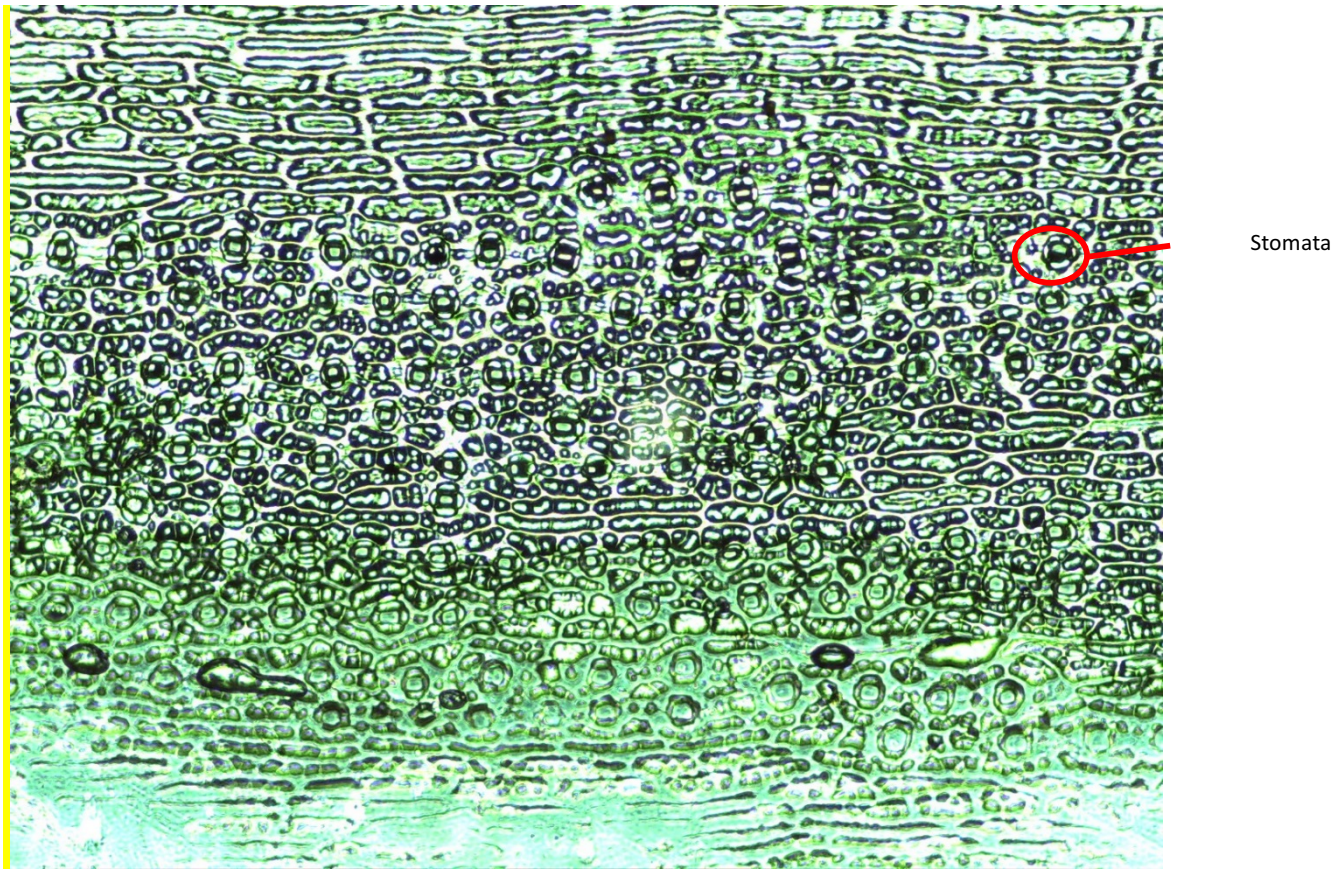
Supplementary Figure 3.3. P16 T. xmedia 'Hicksii' at 20x magnification showing stomata.



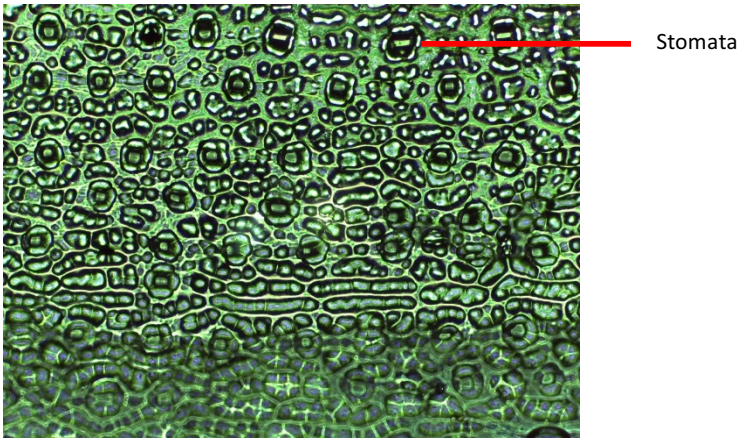
Supplementary Figure 3.4. P14 Taxus brevifolia at 10x magnification. Papillae present on middle of mid rib.



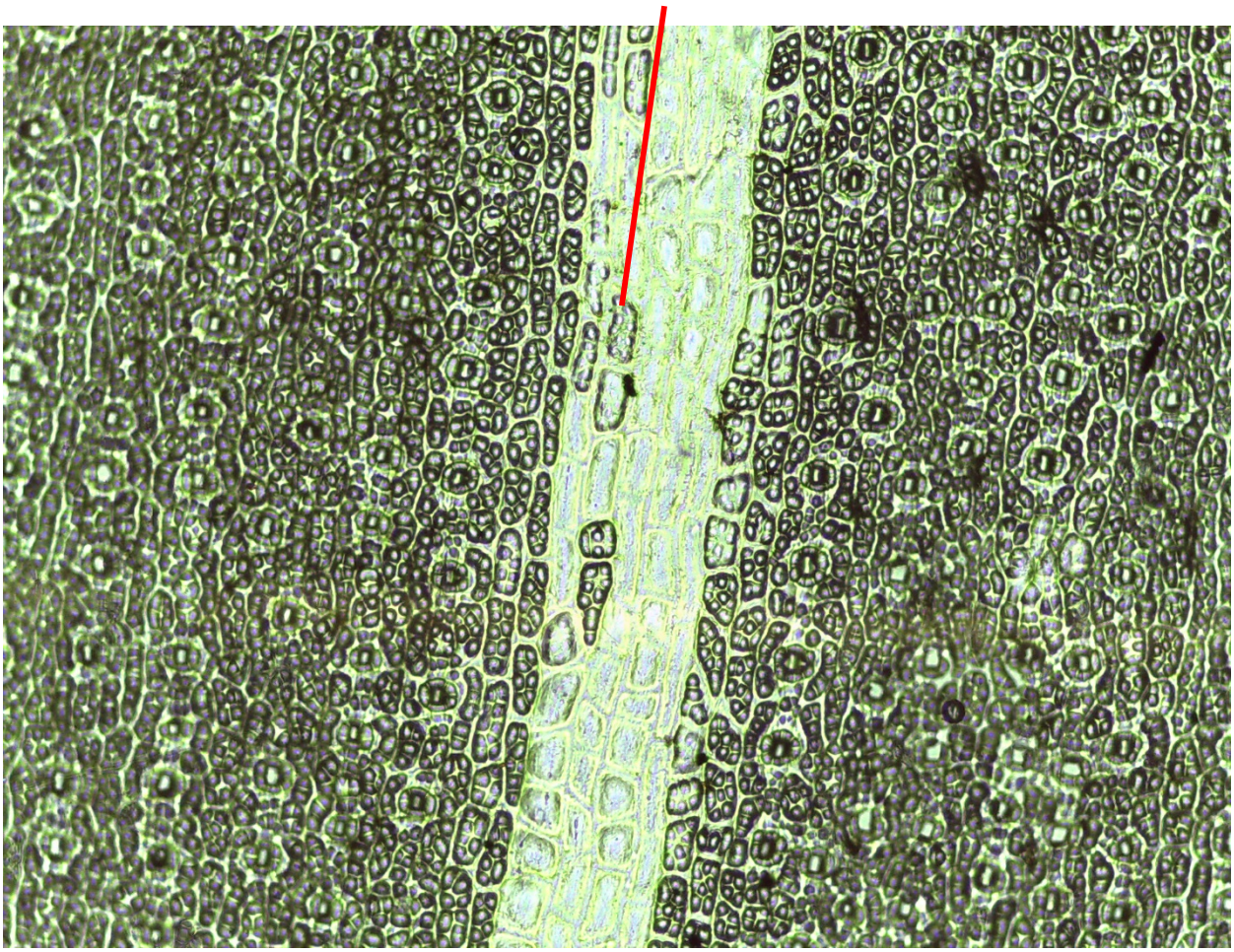
Supplementary Figure 3.5. P16 Taxus x media Hicksii at 10x magnification showing 7 rows of stomata



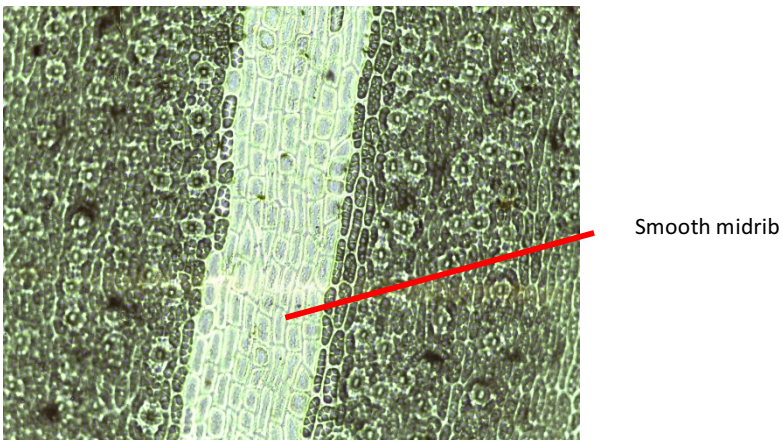
Supplementary Figure 3.6. T1 Taxus baccata at 10x magnification showing 8-10 rows of stomata



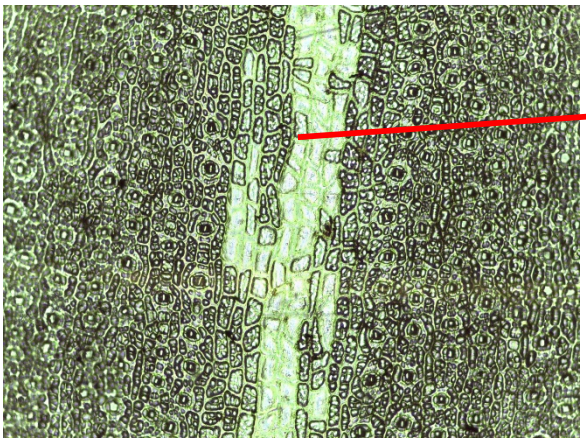
Supplementary Figure 3.7. T1 Taxus baccata 20x magnification showing stomata



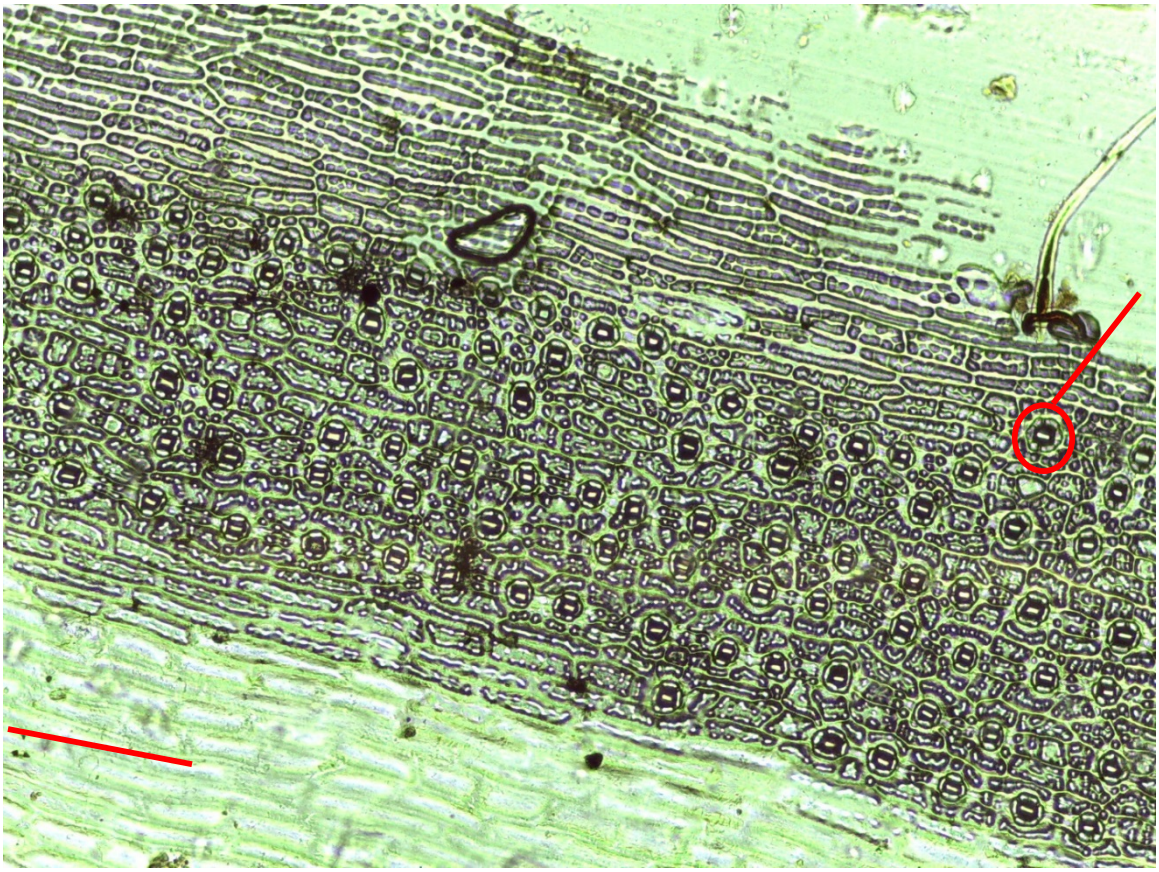
Supplementary Figure 3.8. P29 Taxus baccata "Grandis" at 10x magnification. Some papillae present on the midrib.



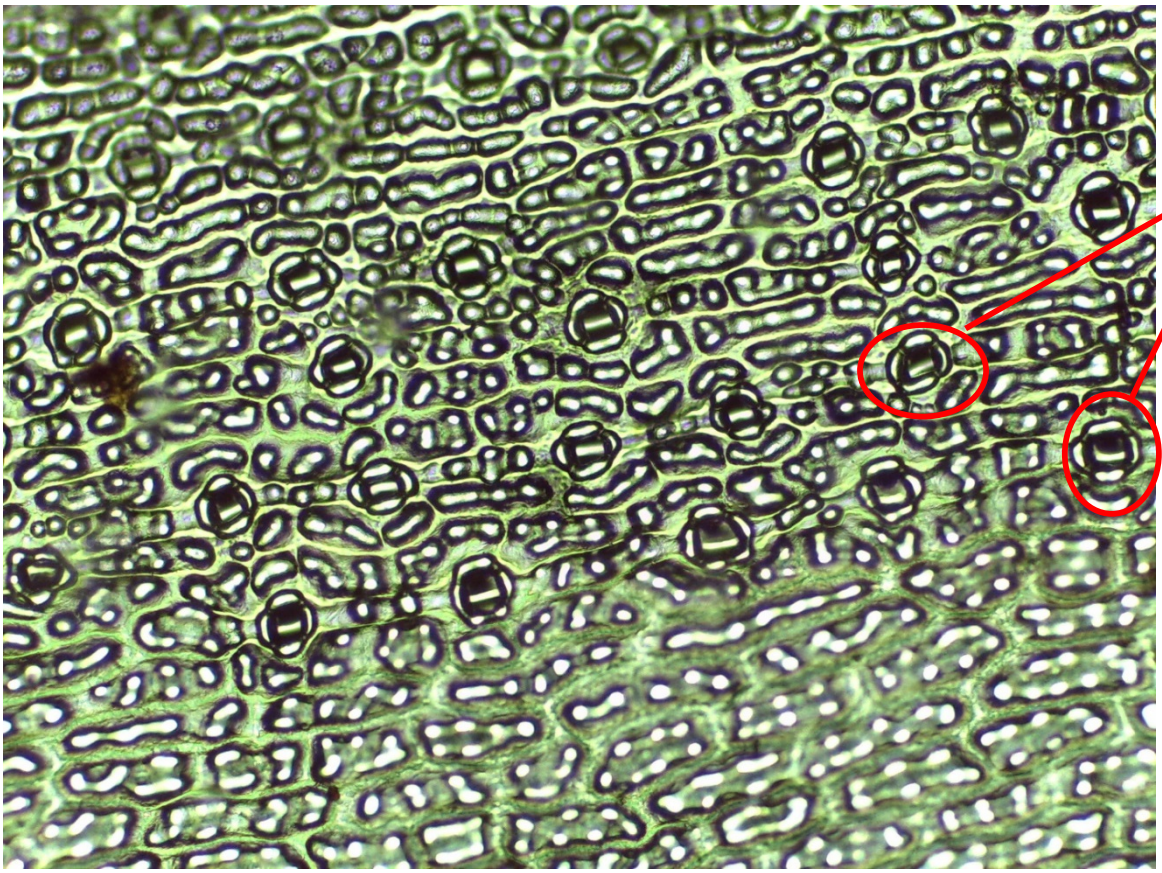
Supplementary Figure 3.9. P29 Taxus baccata "Grandis" 10x papillae absent



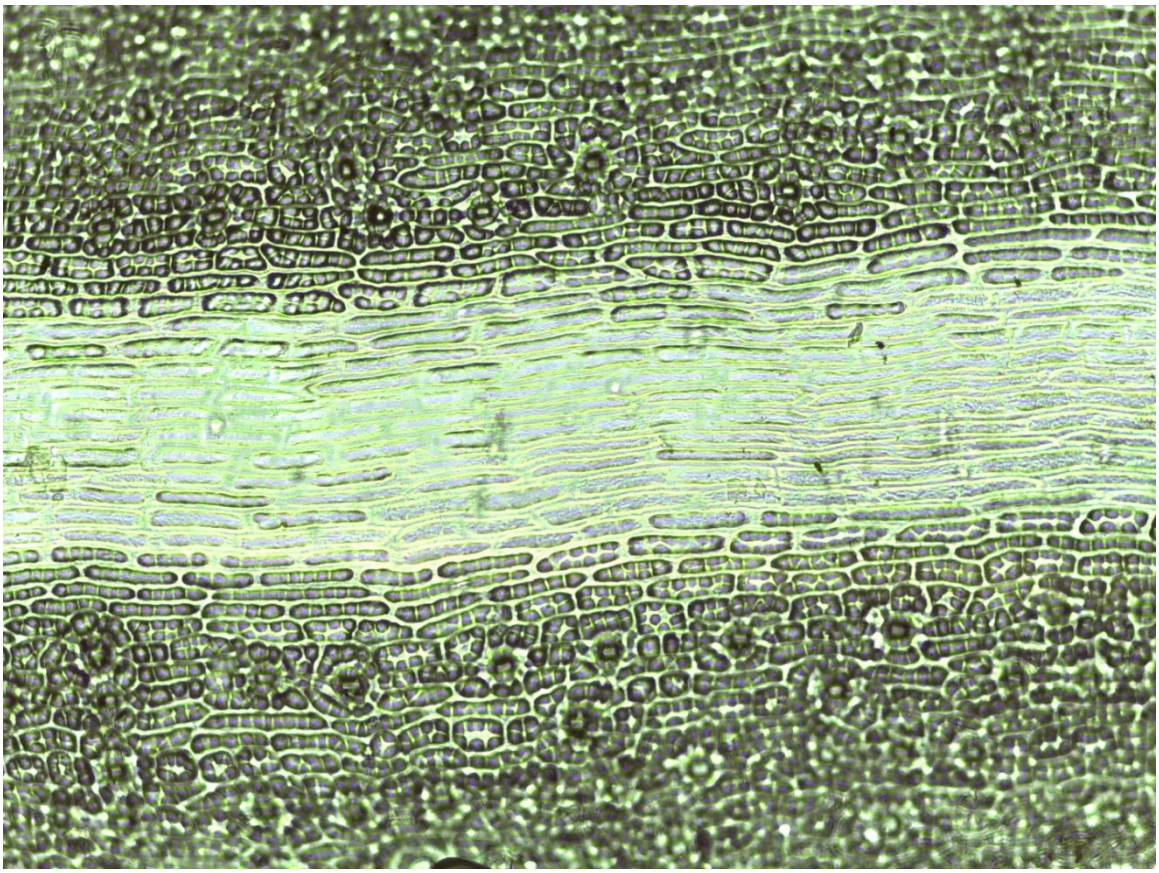
Supplementary Figure 3.10. P29 Taxus baccata "Grandis" 10X papillae on midrib



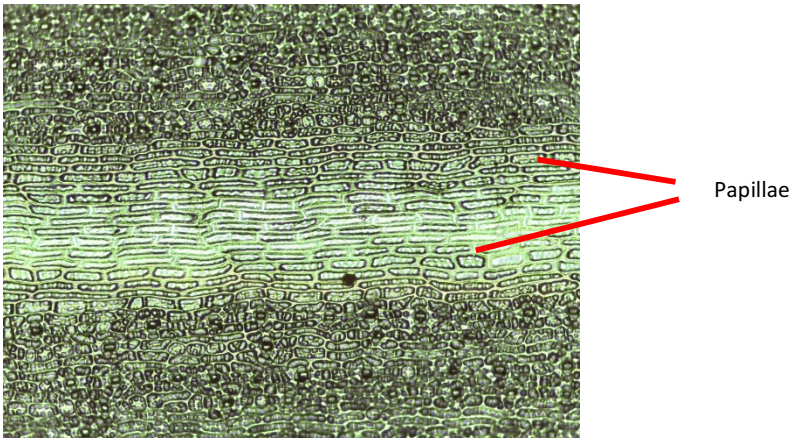
Supplementary Figure 3.11. P11 Taxus canadensis 10x magnification. 5 -6 rows of stomata. Mid rib is smooth in this section of the slide.



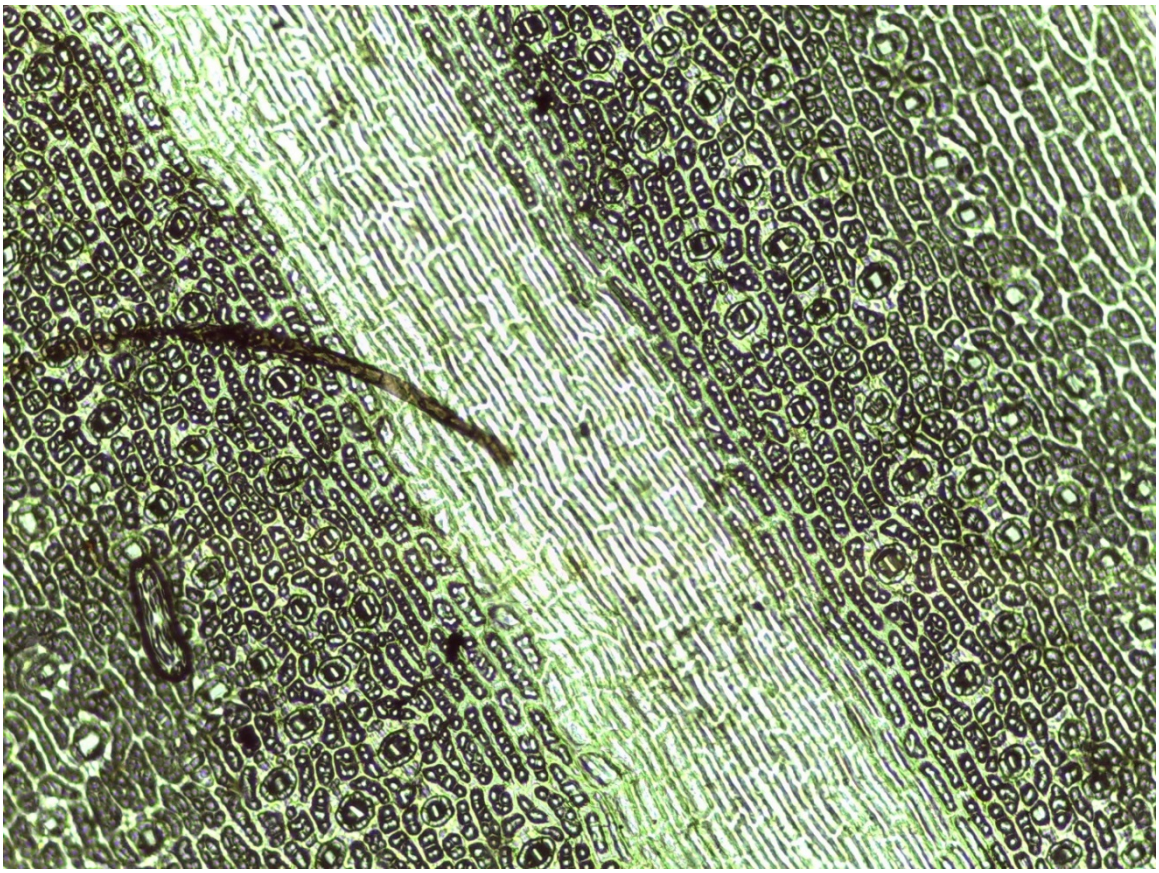
Supplementary Figure 3.12. P12 Taxus cuspidata 'Luteobaccata'. *20X magnification. Stomata.*



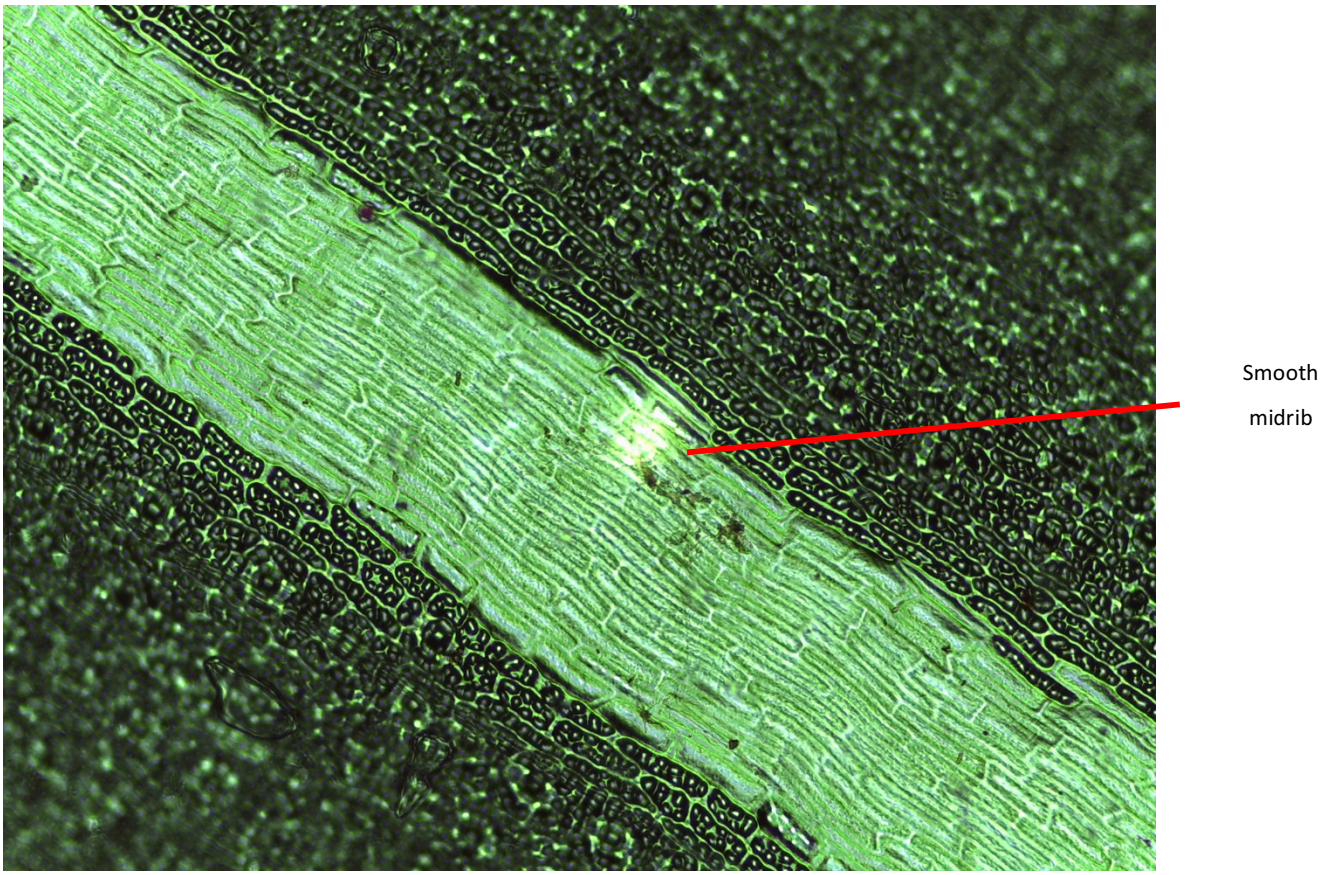
Supplementary Figure 3.13. P17 Taxus cuspidata 10Xmagnification. Papillae absent in this section of leaf mid rib.



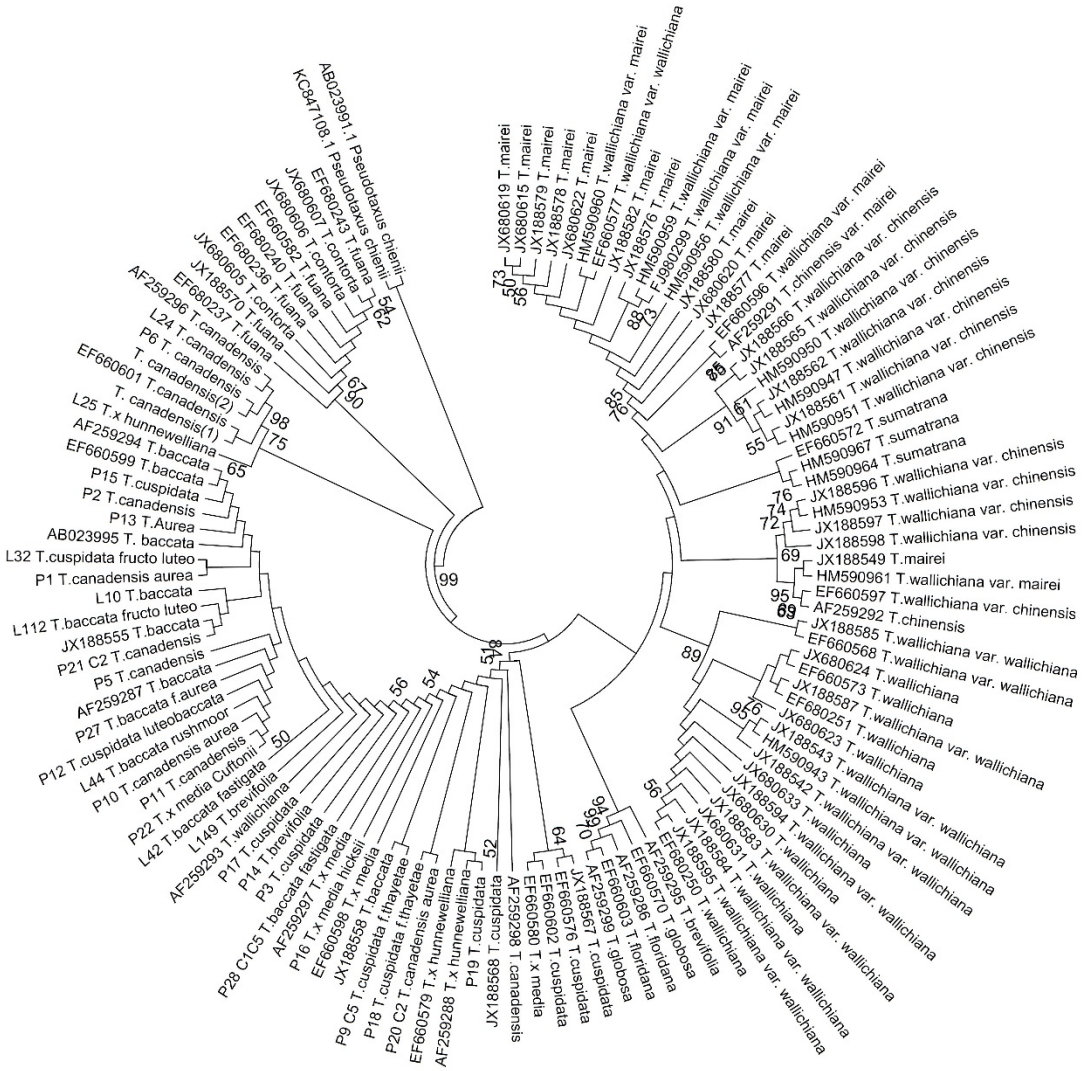
Supplementary Figure 3.14. P17 Taxus cuspidata 10X magnification. Papillae present on this section of midrib



Supplementary Figure 3.15. P14 Taxus brevifolia 10X magnification. Papillae absent near base of leaf on midrib.



Supplementary Figure 3.16. Taxus baccata from Saint Doolagh's Churchyard, Kinsealy, Co.Dublin. Smooth midrib.



Supplementary Figure 3.17 Maximum likelihood bootstrap consensus tree of Taxus ITS sequences showing that the relationship between T. baccata, T. canadensis and T. cuspidata are not well resolved. T. brevifolia is sister to T. globosa and T. flordiana. Taxus fauna groups with T. contorta. Taxus wallichiana is monophyletic but its varieties are not.

Supplementary Figure 4.1. Phylogenetic tree based on ML and Kimura 3 parameter model of TS gene at third codon position and intron/spacer regions (non-coding). The analysis was done to test if a better tree was obtained if only non-coding sites were analysed that were not subject to selection. The tree is not as informative as the analyses based on trnL-F and ITS (Chapter 3). A= mostly T. baccata, T. canadensis and T. cuspidata; B = T. wallichiana, T. fuana and T. sumatrana; C = T. brevifolia, T. floridana

Supplementary table 4.1 Amino acid single letter codes

Name of Amino Acid	**One letter code**
Alanine	A
Arginine	R
Asparagine	N
Aspartic Acid	D
Cysteine	C
Glutamine	Q
Glycine	G
Histidine	H
Isoleucine	I
Leucine	L
Lysine	K
Methionine	M
Phenylalanine	F
Proline	P
Serine	S
Threonine	T
Tryptophan	W
Tyrosine	Y
Valine	V