# Shouldn't we check our data before we date? Some insights from the matrix used by Sauquet et al. (2012)

## INTRODUCTION

Sauquet et al. Sauquet 2012 recently wanted to test the effect of different calibration systems, using fossil or palaeogeographic constraints, secondary dating, etc., for inferences of putative divergence ages within *Nothofagus* based on a dataset of 27 species of this genus, and an nearly complete genus-level sampling of other families in the Fagales. Three taxa were included from possible sister orders of the Fagales: the Cucurbitales and the Fabales. Based on their results, which essentially show that different calibration systems fail to resolve unambiguous divergence ages, they propose a protocol what to do to ensure reliable as possible ages. Their major conclusion is that although models, algorithms, gene- and taxon-sampling may affect the outcome of a dating analyses, the most important issue is the selection of the calibration system. Since their focus was on the latter issue, it appeared to be not relevant to investigate the information content of the signal contained in the used molecular data at the ingroup (*Nothofagus*) and outgroup (other Fagales) level, on which all age estimations are based.

This is hard to understand given the imbalance between both groups: the southern hemispheric *Nothofagus* represents not only a phylogenetically and geographically isolated genus within the Fagales, but also the first diverging branch among the modern Fagales Li 2004. Ecologically the genus is strongly constrained to montane, fully humid, subtropical-temperate climates (Cfa, Cfb, Dfb) which correlates to the overall low interspecific divergence within the four recognised subgenera *Brassospora, Fuscospora, Nothofagus,* and *Lophozonia* Manos 1997 Jordan Hill 1999 Knapp 2005 Acosta Premoli 2010 Mathiasen Premoli 2009 Premoli 2012. Divergence between the subgenera can reach the same level as intergeneric divergences in other Fagales families, but is much lower than interfamily divergences in the other Fagales, which constitute the sister clade of the genus *Nothofagus*

(e.g. Li 2004 Soltis 2011). The remaining Fagales are a genetically, ecologically and morphologically divergent group comprising currently six families APG III. It includes families such as the Fagaceae with genera such as *Fagus,* that go back to the Eocene Manchester Dillhoff 2005 and are strongly divergent from their co-eval siblings (*Quercus* and allies; Paleocene-Eocene; LITINAPPLICATION) already in coding gene regions (data included in Li 2004 Soltis 2011). The Myricaceae and Juglandaceae with equally co-eval first appearances in the fossil record show little or no genetic divergence in the very same gene regions (see data of Li 2004). Fossils considered as "safe" outgroup stem fossils by Sauquet et al. Sauquet 2012, Appendix S2 are of the same Palaeogene age than the used ingroup constraints. Under these pretences, one strongly diverged all-other-Fagales subtree with nodes constrained to be of similar age as their counterparts in the much less diverged *Nothofagus* subtree, it is rather unsurprising that ingroup and outgroup dating schemes cannot converge: markers such as the noncoding spacers of the plastome and nucleome that are needed to resolve the ingroup, are highly divergent in the outgroup, and markers such as coding regions that provide a stable outgroup phylogeny are invariable in the ingroup (see data of Li 2004 Sauquet 2012). As consequence, the ingroup dating using the scheme in Sauquet et al. Sauquet 2012 inflict much too young ages for family-MRCA and generic unfolding within the outgroup; and outgroup dating inflicts relatively young ages for ingroup divergences below the subgenus level that contrast the used ingroup constraints (Table 1; Sauquet 2012 Appendix S4) and the palynological record Dettmann 1990; Hill 1992. Since all inferred ages are minimum ages, they are not necessarily wrong, however, one might ask how informative a reconstructed minimum age of ~10–~40 Ma, is for nodes such as the Betulaceae root node that can be constrained by fossils to be at least ~60 Ma (Sauquet 2012, Appendix S4; Table 1; see also Grimm Renner 2013) or minimum ages of 5–15/10–55 Ma (penalised likelihood/BEAST) for subgenera, which are considered to have diverged prior to the

Gondwana break-up based on fossil evidence Dettmann 1990 Linder Crisp 1995 Ladiges 1998 Svenson 2001.

In addition to fossil constraints Sauquet et al. Sauquet 2012 used tectonic (palaeogeographic) constraints, which is highly problematic in *Nothofagus*: in all subclades a putatively Palaeogene (or earlier) trans-pacific divergence (Table 1) is *followed* by a relatively recent split within areas west of the Pacific, which contrasts vicariant scenarios for the unfolding of the subgenera linked to the (East) Gondwana break-up (e.g. Svenson 2001; Knapp 2005). A vicariant scenario would predict first an isolation of New Caledonia-New Zealand from Australia/South America before 55.8 Ma (or 80 Ma, Svenson 2001), the latter two remaining connected via Antarctica until 30–40 Ma Svenson 2001 Sauquet 2012. The genus persisted in Antarctica at least until the Miocene (c. 14 Ma)/?Pliocene (3 Ma; Hill 1996 Fleming Barron 1996; but see Ashworth et al. abstract, American Geophys. Union, 2009, http://adsabs.harvard.edu/abs/2009AGUFMPP41D..06A), with some species being resilient to relative harsh, subarctic climates Francis Hill 1996. It is unknown if *Nothofagus* may be dispersed by migrating birds, but the overall low trans-oceanic genetic divergences and high morphological similarities between Australian and New Zealand species have led some authors to assume long-distance dispersal for the genus (e.g. Hill 1992 Knapp 2005). Alternatively, short-distance dispersal from extinct populations should be considered. The pollen record of all four subgenera of *Nothofagus* can be traced back to the late Cretaceous of Antarctica; and micro- and macrofossils evidence the existence of subgenera *Nothofagus* and *Brassospora* far outside their modern range (see Svenson 2001, table 2) in the late Cretaceaous and Palaeogene. Ancient or subrecent reticulation may have altered genetic signal. For instance, the plastomes of modern members of subgenus *Nothofagus* are not species-specific but geographically controlled Acosta Premoli 2010 2012, which could indicate that (i) the modern species are the product of widespread introgression with already isolated lineages that escaped from Antarctica into South America or (ii) that chloroplast

capture is the rule rather the exception in *Nothofagus,* hence any plastid-based inference cannot be used straightforwardly to estimate species divergence times at all. Antarctica hosted populations of *Nothofagus* until the mid-Pliocene facing Australia (c. 3 Ma; Fleming Barron 1996) and South America LIT. Single colonisation events from Antarctica/Australia via New Zealand, ≥ 40 Ma, alternatively via a (hypothetical) New Guinea-New Caledonia island-arc system, 8–10 Ma, are the most parsimonious explanation for the existence of modern New Guinean (not known before 5–7 Ma; Dettmann 1990 Hope 1996) and New Caledonian (Quarternary record only, Dettmann 1990) members of *Brassospora* Svenson 2001. Svenson et al. Svenson 2001 also concluded that the New Zealand species *N. gunnii* (subgenus *Lophozonia*) and *N. menziesii* (subgenus *Fuscospora*) could only be long-distance dispersed, in the case of *Lophozonia* the dispersal was assumed to have taken place 30 Ma *after* the formation of the Tasman Sea and the isolation of New Zealand from Australia-Antarctica-South America. All three effective ingroup palaeogeographic (vicariant) constraints used by Sauquet et al. Sauquet 2012 relate lineages/divergence, for which Svenson et al. Svenson 2001 and earlier studies predicted long-distance dispersal or recent short-distance dispersal via an island-arc as the *only* possible explanation. In conclusion, already the conception of the paleogeographic scenario used in Sauquet et al.'s Sauquet 2012 is fundamentally flawed, hence, its failure to reconstruct meaningful ages to be expected.

All dating programs either require a fixed input topology, a tree with defined branch lengths (r8tes, …) or optimize simultaneously the input topology, branch-lengths, and subsequently the divergence ages (BEAST; MrBayes). Hence, one either has to ensure the stability of the tree topology or to calculate divergence ages using different topological scenarios. The inferred branch-lengths should be meaningful regarding the divergence between taxa. Both issues rely heavily on the amount of accessible data. For their evaluation of different calibration systems, Sauquet et al. Sauquet 2012 relied on a concatenated nucleotide matrix with 5444 nucleotide positions (sites) comprising the plastid genes *atp*B

and *rbc*L (representing 51% of total sites), the intergenic plastid *atp*B-*rbc*L and *trn*L-*trn*F spacers, the plastid *trn*L intron (27%), and the nuclear-encoded ITS region (12%). According to a NCBI GenBank extract, these markers cover 52% of the currently available sequence data: in total 216 accessions are available, of which 93 were included in the study. Two further plastid regions, the region comprising the *psb*B, *psb*T, *psb*N, and *psb*H genes (*psb*BTNH complex) and the *trn*H-*psb*A spacer Acosta Premoli 2010 Mathiasen Premoli 2009 were not included (a total of additional 106 sequences of all subgenera except for *Brassospora*). In general, > 15 year old sequences with a number of potential sequencing and editing artefacts were preferred over newer data from the according species; in particular this is true for the most variable regions within the ingroup of *Nothofagus* (*atp*B-*rbc*L, ITS).

Another shortcoming is that the study lacks basic information that would be needed to assess the signal behind the inferred topologies; the 'gappyness' of the matrix, the levels of ingroup and outgroup divergence in coding and non-coding plastid regions and the ITS, and if signal from the nuclear ITS is fairly compatible with that of the plastid regions, at least for the ingroup. Sauquet et al. Sauquet 2012 note that earlier dating studies resulted in extremely young ages, but do not discuss the generally low divergence below the subgeneric level in *Nothofagus* (note the relatively large BEAST confidence intervals even if more than 50% of the nodes are constrained in the study of Sauquet 2012). ML has been shown to be relatively stable against missing data regarding the topology of a tree, but it affects naturally the estimation of branch-lengths LIT, hence, the estimated heights of nodes in chronograms. Furthermore, the authors opt to exclude a number of Fagales genera (*Alnus, Castanopsis, Corylus, Pterocarya,* divergent family of the tropical-subtropical Casuarinaceae, with three to four currently accepted genera, only represented by one taxon), for which they compiled "safe" constraints. The placeholder species used for two of the Juglandaceae genera (*Alfaropsis, Engelhardia*) refer to the same species (*A. roxburghiana* ≡ *Engelhardia roxburghiana,* and *E. fenzelii,* a synonym of the latter; Flora of China 1999).

Last but not least, investigating congruence between nuclear and plastid genealogies is obligatory at the genus- and species-level. Plastid haplotypes and ITS variants are propagated by two fundamentally different pathways in angiosperms: uniparentally in the case of the plastome, whereas the nucleome is inherited from both parents, which, in the case of plants means that there is always the possibility of long-distance genetic exchange via pollen, independent of the general dispersal mode of the plant. For the ingroup, the only comprehensive studies at the subgneric-species level demonstrate that ITS and plastid data differ fundamentally in their signal (Acosta Premoli; subgenus *Nothofagus*). For the Betulaceae and Fagaceae, plastid and ITS data show not necessarily strongly conflicting signal Grimm Renner 2013, but different preferences for topological alternatives (discussed in Forest 2005; Denk Grimm 2010). Ribeiro et al. Ribeiro 2011 found that there are at least two probably functional, non-orthologous NORs in some of the Fagaceae, which may explain the high levels of intra-individual ITS variability found in *Fagus* and *Quercus* Denk 2002 Denk 2005 Denk Grimm 2010, a phenomenon commonly addressed as 'ITS paralogy' in literature. Similarly sampled data is so far only available for the five species of subgenus *Nothofagus* Acosta Premoli. If the nuclear and plastid partitions prefer different topologies, this could indicate that the plastome and nucleome have different evolutionary histories, thus, their genealogies may not be straightforwardly translated into a species-phylogeny. Furthermore, a tree based on such a matrix, can be expected to have too long branch-lengths (Fig. 1). Intra-species variation due to reticulation process may further obscure inter-species divergences. In case of *Nothofagus* subgenus *Nothofagus* repeated chloroplast capture has also been invoked as explanation for the distinct, geographically linked plastid haplotypes shared by all species, including coding and non-coding regions Acosta Premoli 2010 Premoli 2012. An additional explanation is that *Nothofagus* tends to retain ancient plastid polymorphism at the species level; e.g. induced by widespread introgression/ hybridisation predating the formation of the modern.

By subjecting the data matrix used by Sauquet et al. Sauquet 2012 to a comprehensive phylogenetic re-investigation, the relevance of the original study regarding the comparison of various calibration systems is questioned. The used data does not allow concluding on a single unanimous tree, which renders the assignation of fossil constraints as most-recent common ancestors to distinct nodes of the tree preferred by the concatenated data problematic. The amount of missing data and selection of markers can hardly provide meaningful branch-lengths, neither for the focus group ('ingroup') *Nothofagus* nor to compare this group with other Fagales (as 'outgroup'); accordingly the comparison of 'ingroup' and 'outgroup' dating-derived ages is meaningless *per se* and explains the huge discrepancy in the inferred node heights. Thorough investigation of the data stored in gene banks confirms that species of subgenera of *Nothofagus* cannot be unanimously distinguished based on available plastid data; and that identical or highly similar plastid haplotypes can be found in individuals of the same subgenus but of different provenances (such as New Zealand and South America), which is in strong contrast to reconstructed and constrained age estimates. Furthermore, the assumed "safety" of used minimum age constraints is critically addressed. It is recommendable that studies using empirical data of various sources to address theoretical and empirical problems meet highest possible standards on both ends and should not exclusively rely on unrepresentative data, incomprehensive phylogenetic hypotheses and simplistic conceptions of evolution. An outline is provided what future studies should investigate in order to make molecular-based analyses beneficial for a deeper understanding of *Nothofagus* and, in general, the Fagales.

**TODO Figure 1.** Branch-lengths in separate and concatenated trees.

**Table 1 (following page).** Estimated heights (divergence ages) of nodes for which "safe" constraints were used.

Table 1

| Node | Clade | Fossil constraint | Tectonic constraint | "Safe" ingroup constraints | | "Safe" outgroup constraints | | Palaeogeographic constraints | |
|---|---|---|---|---|---|---|---|---|---|
| | | | | ML–PL | BEAST | ML–PL | BEAST | ML–PL | BEAST |
| A | Fagales | ≥ 83.5* | [≥ 55.8] | 84.5–103.6 | 90.8–123.7 | 101.0–114.0 | 112.5–124.8 | 99.9–113.6 | 105.5–124.7 |
| L | Nothofagus | [≥ 31.5] | [≥ 55.8] | 41.1–47.7 | 43.9–81.2 | 24.7–39.3 | 44.5–95.5 | 70.0–78.2 | 87.4–113.6 |
| M | subgenus Lophozonia | ≥ 31.5 | ≥ 40 [≥ 55.8] | 31.5 | 31.5–46.6 | 7.9–14.6 | 12.7–53.0 | 56.8–60.4 | 59.3–90.6 |
| V | Western South Pacific *Lophozonia* | | ≥ 55.8 | 4.9–10.7 | 4.6–27.2 | 4.0–9.0 | 4.1–30.9 | 55.8 | 55.8–73.2 |
| N | N. cunninghamii+N. moorei | ≥ 0.8 | | 2.5–6.3 | 1.2–15.2 | 2.2–6.0 | 1.4–18.8 | 6.4–17.3 | 2.6–53.4 |
| O | subgenus Fuscospora+subgenu | [≥ 31.5] | [≥ 55.8] | 32.4–35.7 | 34.0–55.8 | 14.1–24.7 | 28.3–72.1 | 63.0–69.5 | 72.4–103.9 |
| W | subgenus Fuscospora | [≥ 55.8] | ≥ 40 [≥ 55.8] | 7.1–13.8 | 9.2–37.0 | 7.1–15.0 | 10.0–48.5 | 57.4–62.4 | 59.0–88.0 |
| P | Western South Pacific *Fuscospora* | | ≥ 55.8 | 3.2–7.0 | 3.3–20.3 | 3.4–8.4 | 4.3–29.0 | 55.8 | 55.8–72.3 |
| Q | subgenus Nothofagus+subgenu | ≥ 31.5 | ≥ 40 [≥ 55.8] | 31.5 | 31.5–44.1 | 11.1–20.0 | 19.4–56.7 | 59.5–65.3 | 60.1–89.6 |
| T | subgenus Nothofagus | | | 6.1–11.6 | 7.6–27.3 | 5.8–11.9 | 8.8–37.3 | 11.5–24.7 | 12.6–59.8 |
| X | subgenus Brassospora | | ≥ 55.8 | 5.5–12.4 | 10.3–28.1 | 4.7–10.6 | 11.7–37.8 | 55.8 | 55.8–70.2 |
| R | New Guinean *Brassospora* | | | 2.9–7.6 | 5.1–22.0 | 3.0–8.4 | 6.7–30.0 | 6.0–19.2 | 9.4–60.9 |
| B | Other Fagales | ≥ 83.5 | | 66.8–83.1 | 70.4–110.8 | 92.6–103.5 | 100.2–118.5 | 83.2–97.9 | 85.8–117.6 |
| C | Fagaceae | ≥ 47.0 [≥ 64.4] | | 49.5–64.2 | 36.1–86.1 | 75.6–87.8 | 66.4–102.0 | 62.5–78.3 | 41.3–100.2 |
| D | Quercoideae | ≥ 37.2 [≥ 43.8] | | 12.2–18.4 | 15.0–43.1 | 37.2–48.1 | 43.8–66.9 | 16.8–24.7 | 22.6–61.0 |
| E | Castanea+Lithocarpus+Chrysol | ≥ 43.8 | | 10.0–15.2 | 10.6–33.3 | 43.8–43.8 | 43.8–57.0 | 13.9–20.7 | 14.0–49.1 |
| S | core Fagales | [≥ 64.4] | | 54.6–67.3 | 53.4–94.8 | 86.2–95.6 | 90.0–110.2 | 70.9–84.1 | 69.7–106.6 |
| F | Casuarinaceae+Ticodendraceae | ≥ 55.8 | | 48.6–61.0 | 37.1–79.1 | 78.9–89.3 | 74.9–99.4 | 64.1–77.4 | 40.3–88.5 |
| G | Betulaceae | ≥ 59.8 | | 20.7–30.2 | 7.5–41.6 | 59.8 | 59.8–71.3 | 28.0–40.8 | 6.2–47.7 |
| H | Rhoipteleaceae+Juglandaceae | | | 31.6–42.1 | 33.6–73.9 | 73.3–80.5 | 72.5–96.0 | 42.8–56.1 | 46.8–88.3 |
| I | Juglandaceae [s. APG II] | ≥ 64.4 | | 18.8–25.8 | 22.3–52.2 | 64.4–64.7 | 64.4–80.3 | 25.8–35.4 | 31.7–73.1 |
| J | Juglandoideae | ≥ 52.7 | | 13.2–19.2 | 12.3–37.4 | 57.7–62.3 | 58.4–72.7 | 18.6–26.9 | 18.0–55.8 |
| K | Juglandinae | ≥ 55.8 | | 6.6–12.0 | 4.3–22.9 | 55.8 | 55.8–63.5 | 9.3–17.0 | 3.3–33.6 |

* Although outgroup, included in the ingroup constraint analysis

Colour code for age constraints cf. Sauquet et al. 2012

Blue = effective age estimates for this node were directly affected by this constraint

Black = ineffective constraint, e.g., node much older than minimum age constraint applied

Grey = superfluous, since an equal or older minimum age constraint was used above this node; i.e., might as well be deleted

[red] = implied by an effective age constraint at a connected node

Too young inferred ages according "safe" constraints used in other analyses

Inferred age more than 50% older than "safe" constraint at corresponding nodes

## MATERIAL AND METHODS

The basic data matrix used here is exactly the same used in Sauquet et al. Sauquet 2012 dryad link, and the data has been handled in exact the same fashion than in the original study regarding the partition of the data. To illustrate gene-sampling issues regarding problematic intergeneric relationships and estimation of branch-lengths in the Fagales as a whole, the matrix of Li et al. Li 2004 was used; stored data from *Nothofagus* was revisited (18S rDNA, nuclear rRNA gene: 2 accessions from 2 subgenera; *atp*B, plastid gene: 12/3; *mat*K, plastid gene: 5/3; *mat*R, mitochondrial gene: 2/2; *trn*L region, plastid intron and spacer: 11/3; *rbc*L, plastid gene: 48??/4). In addition, the global gene banks were mined for all available data on *Nothofagus* (downloaded 20/03/2012), providing up to six partitions that cover some intra- and inter-specific variation: the nuclear ITS region of the 35S rDNA cistron; and the plastid *trn*H-*psb*A, *trn*L-*trn*F intergenic spacers, the *psb*BTNH gene complex, and the *rbc*L gene and adjacent *atp*B-*rbc*L spacer. For comparison, data of the three principal gene regions used in Sauquet et al. Sauquet 2012, i.e. the complement comprising the *atp*B and *rbc*L genes and their intergenic spacer, the complement comprising the *trn*L gene, intron, and *trn*L-*trn*F spacer, and the ITS region of the 35S rDNA, of the Fagaceae was mined (downloaded 22/3/2012) using the following search strings: "rbcL [gene] OR atpB [gene] OR atpB-rbcL OR atpB-rbcL"; (2) "trnL [gene] OR trnL-trnF" {For *Lithocarpus* "(trnL [gene] OR trnL-trnF OR trnL-F) NOT matK [gene] NOT 5.8S"}; (3) "5.8S" [This search string downloads also five short, pseudogenous ITS2 fragments of *Quercus* Group *Cyclobalanopsis* (T.A. Ishida, S.G. Goto, H. Sato, M.T. Kimura, unpublished 05/09/2000) and three putative *Carya* ITS (Juglandaceae) from environmental samples. One environmental sequence is of very poor quality and not considered (EU646157)].

As far as new data was used, sequence alignments and initial tree analyses relied on MAFFT Katoh 2005. The standard setting for fast-alignment generation was used. Auto-generated alignments were checked visually for consistency in Mesquite Maddison Maddison; see

corresponding NEXUS files for adjustments and modifications of standard procedure. All alignment files used in the current study are archived and available for download at www.palaeogrimm.org/data (see HowToHandleFiles.txt for user instructions).

Final ML and bootstrap analyses used a RAxML v. 7.2.6 Stamatakis 2006 Stamatakis 2008 with the following settings: xxxx. In contrast to the original study, not only the concatenated data was analysed but also subsets of the data: single partitions, data sets excluding one of the partitions, and data sets including only the combined coding or non-coding plastid regions.

The tree-likeliness of each matrix, and individual taxa in the matrix, was estimated using Delta values Holland Moulton, which can, among other statistics, be calculated with the executable Dist_stats devised by M. Göker Göker. General differentiation patterns are illustrated using planar phylogenetic networks, which, in contrast to trees, can handle incompatible signal inherent to many data sets focussing at or below the genus level. Phylogenetic networks were computed using uncorrected *p*- and ML-based distances using the neighbour-joining algorithm Bryant Moulton implemented in SplitsTree 4 Huson Bryant 2006. ML-based distances were obtained using the parameters as optimised by RAxML under a GTR+Γ model and with different partitions for each sequence region, and, in case of plastid genes *atpB* and *rbcL*, the 1st, 2nd and 3rd codon position. Maximal intra- and any minimal inter-taxon distances were computed using the program PBC devised by M. Göker Göker Grimm 2008. The consensus network approach Holland was used to illustrate topological alternatives in collections of trees ('best-known' ML and BS replicate trees inferred with RAxML).

## RESULTS & DISCUSSION

### *Matrix composition*

About a third of the characters in the original concatenated matrix (WG32MS) are either gaps or missing data (Table 2). The matrix includes several taxon pairs with no or limited

sequential overlap; in three of the six partitions, more than one third of the taxa are completely missing, and 56–71% of the focus group (genus *Nothofagus*). The most underrepresented group is that of *Nothofagus* subgenus *Brassospora,* with 23 accessions (out of 60 possible) covering three out of the six partitions, the *rbcL* gene (complete), the *atpB-rbcL* spacer (3 missing), and the ITS (five missing, only old data used, including a contaminant or mis-labelled sequence of one species[1]). Of the other families of the Fagales, the Betulaceae are the most under-represented, four of the six currently accepted genera are missing. Due to confusion inflicted by synonymy, no data was included from *Engelhardia* s.str. (Juglandaceae: Engelhardioideae), the selected taxon, a chimera between *E. fenzelii* and *E. roxburghiana* (cf. Sauquet 2012, Appendix S1; no ITS data included) is conspecific to *Alfaropsis (= Engelhardia) roxburghiana* (3 out of 6 partitions missing; cf. Flora China 1999). Within the three non-Fagales, two to four partitions are not represented. The matrix Delta value (mDV) of the concatenated matrix is relatively high (mDV = 0.32), and much higher than that of single-gene data sets, a first evidence that the signal from the various partitions in the matrix is not compatible with each other. Signals from matrices with *good* taxon coverage are *less* tree-like than the signal from matrices with many missing accessions (Table 2). Due to the significant amount of missing data in some partitions, i.e. differing taxon-samples, the matrix (mDV) and individual Delta values (iDV) based on the complete taxon set (Table 2) cannot be straightforwardly compared. If the data is reduced to only those taxa that are represented in all partitions (23 taxa, including 12 *Nothofagus* species of three of the four subgenera, the other species coming from the Betulaceae, Fagaceae, and

[1] The old ITS sequence data used in Sauquet et al. Sauquet 2012 show a number of derivations compared to more recent sequences; the most obvious is the 31 nt-long "deletion" in the central part of the 5.8S rDNA, only found in the first set of ITS sequences reported by Manos et al. (Manos 1997; accession numbers U96849–U96870). In contrast to more recent sequences, the old sequence reported for *N. dombeyi* shows the variant of the closely related, but ITS-distinct *N. pumilo.* Old sequencing facilities had a higher error rate and produced shorter reads, and errors such as contamination/mis-labelling had a higher risk to remain undetected (see Grimm 2003 for examples from *Acer* and *Fagus*). Old sequence data should therefore always be compared, replaced or complemented with newer data if available to minimise such errors. [But this is, however, not a standard in plant phylogenetic analyses, see e.g. Soltis 2012, who included artificial sequence based on nearly 20 year-old back-transcribed rRNA fragments in their data instead of complete 18S rDNA sequences, e.g. used by Soltis 2004 Soltis 2007, available for the same taxa]

Juglandaceae), the most tree-like signal comes from the *atp*B partition, and the least tree-like signal from the *atp*B-*rbc*L spacer and the ITS.

The taxa with the least tree-like signals in the concatenated matrix are several species of *Nothofagus* (three of subgenus *Brassospora*; one of subgenus *Fuscospora*) that lack three and four of the six partitions and those two outgroup taxa (2 and 3 partitions missing) which have no mutual sequence overlap with three accessions of the Juglandaceae: Engelhardioideae (3 partitions missing; see Table 3 for a summary). A clear trend is visible for the concatenated data: taxa with a lot of missing data provide much less tree-like signal than those with no missing data (Table 3). In other words, the partitions represented in these taxa do not have the power to compensate for missing signal of the missing partitions. This agrees also with the general range of iDV of well represented taxa (lacking not more than one partition) based on the concatenated data, which are relatively high compared to other multigene data sets (Table 4). If the signal from the combined data would be additive and largely compatible, one could expect that mDV of concatenated matrices decrease, because of general coalescence.

**Table 2.** Some matrix statistics.

**Table 3.** Some taxon statistics.

**TODO Table 4.** Some mDV and iDV ranges of multigene datasets.

## *The nature of the signal in the concatenated data*

Figure 2 demonstrates the basic problem of any phylogenetic tree inference or molecular clock optimisation with the concatenated data used by Sauquet et al. Sauquet 2012: the data contains limited information regarding the phylogenetic backbone of the Fagales and relationships within terminal groups. Relatively faint differentiation patterns correlate to major divergences, (i) the unfolding of the major lineages within the Fagales and (ii) the diversification of *Nothofagus* supposedly linked to the final Gondwana break-up in the latest Cretaceous and early Palaeogene, according to the age constraints imposed by Sauquet et al.

Table 2

| Matrix | Number of taxon pairs with no sequential overlap | In percent | Gappyness* | Proportion of taxa without data | Only *Nothofagus* | Number of distinct alignment patterns† | Only *Nothofagus* | Matrix DV (all taxa) | Matrix DV (only taxa with complete data) |
|---|---|---|---|---|---|---|---|---|---|
| Concatenated data | 6 | 0.2% | 35% | 0% | 0% | 1494 | | 0.317 | 0.064 |
| - *atpB* gene | | | 33% | | | 1695 | | | |
| - *rbcL* gene | | | 42% | | | 1518 | | | |
| - *rbcL-atpB* spacer | | | 34% | | | 1494 | | | |
| - *trnL-trnF* spacer | | | 34% | | | 1650 | | | |
| - *trnL* intron | | | 33% | | | 1640 | | | |
| - ITS | | | 35% | | | 1608 | | | |
| *atpB* gene | 810 | 31.1% | | 39% | 56% | 51/28/147 | | 0.073 | 0.061 |
| *rbcL* gene | 147 | 5.7% | | 6% | 0% | 94/74/235 | | 0.281 | 0.113 |
| Coding cpDNA | 147 | 5.7% | | 6% | 0% | 629 | | 0.273 | 0.160 |
| *rbcL-atpB* spacer | 455 | 17.5% | | 20% | 11% | 427 | | 0.155 | 0.117 |
| *trnL-trnF* spacer | 747 | 28.7% | | 37% | 71% | 271 | | 0.112 | 0.098 |
| *trnL* intron | 810 | 31.1% | | 39% | 56% | 281 | | 0.121 | 0.085 |
| Non-coding cpDNA | 247 | 9.5% | | 8% | 11% | 552 | | 0.303 | 0.084 |
| ITS | 414 | 15.9% | | 12% | 19% | 313 | | 0.193 | 0.152 |

* Proportion of gaps and completely undetermined characters

† In plastid genes, shown separatedly for 1st, 2nd, and 3rd codon position.

‡ BS search did not converge

Table 3

| | Number of taxa | Number of missing partitions | Range of iDV, concatenated matrix |
|---|---|---|---|
| Outgroups | 2 | 1–3 | 0.36–0.42 |
| Betulaceae | 2 | 0 | 0.26/0.27 |
| *Casuarina* (Casuarinaceae) | 1 | 1 | 0.32 |
| *Chrysolepis* (Fagaceae) | 1 | 3 | 0.35 |
| Other Fagaceae | 5 | 0 | 0.25–0.29 |
| Juglandaceae | 9 | 0–3 | 0.24–0.40 |
| Myricaceae | 2 | 1 | 0.30/0.33 |
| *Nothofagus* | | | |
| Subgenus *Brassospora* | 10 | 3–4 | 0.31–0.44 |
| *Nothofagus cliffortioides* (sg. *Fuscospora*) | 1 | 4 | 0.45 |
| Subgenus *Fuscospora*; others | 5 | 0 | 0.27–0.29 |
| *Nothofagus alpina* (sg. *Lophozonia*) | 1 | 3 | 0.33 |
| Subgenus *Lophozonia*; others | 5 | 0 | 0.27–0.28 |
| Subgenus *Nothofagus* | 5 | 0/3 | 0.29–0.32 |

Sauquet 2012. Some parts of the graph are dominated by box-like structures, indicating that, despite the found support for alternative relationships (Fig. 2), the signal in the concatenated data may still be incompatible.

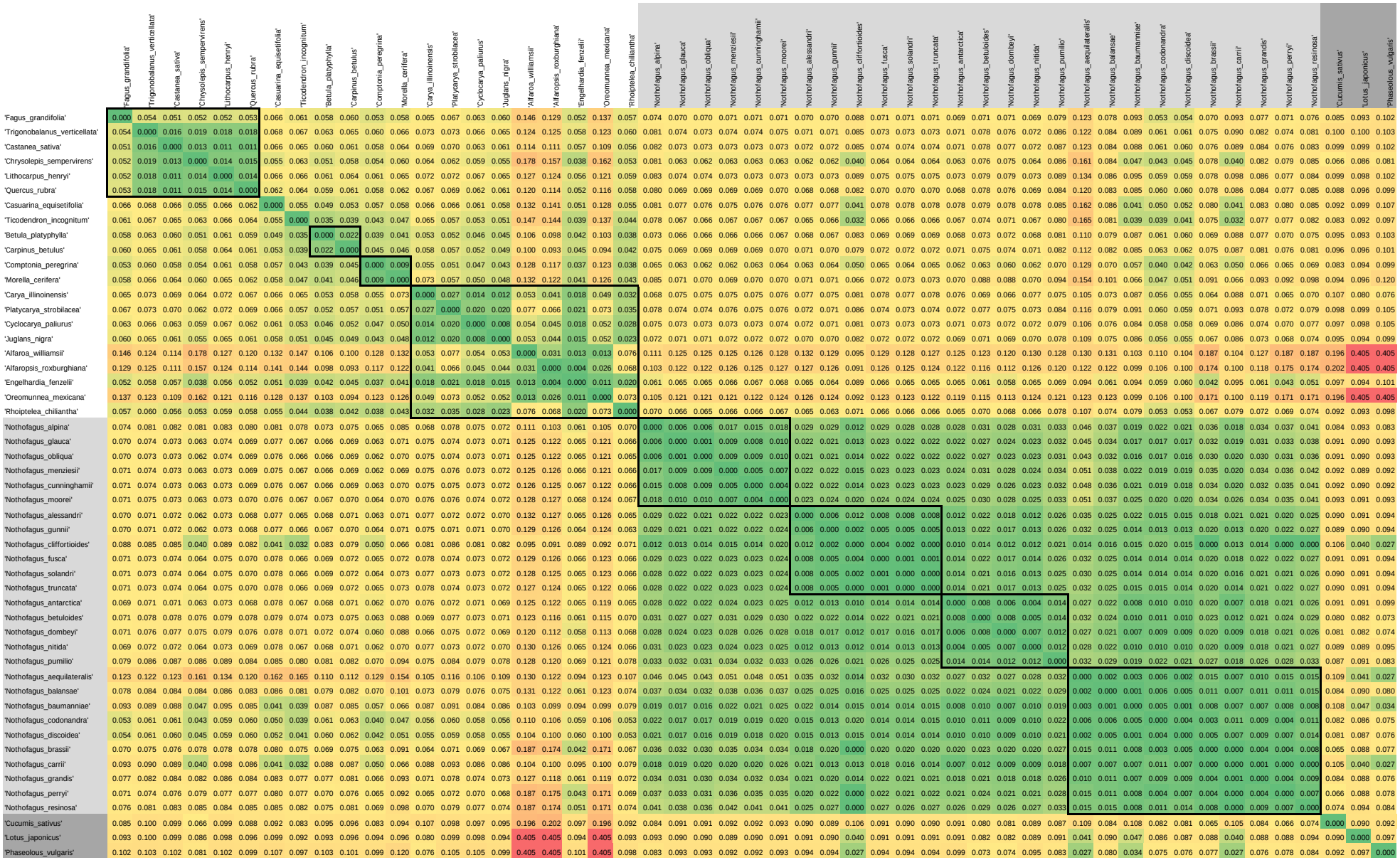
There are two reasons for decreased bootstrap support: low, insignificant differentiation in the data REF and internal conflict, signal incompatibility (e.g. Zander 2004). A ML BS support ($BS_{ML}$) of 60 in the optimal case may indicate that 60% of the site patterns support this split, and the remaining 40% are indecisive. It may also indicate that 40% of the variable sites prefer a different split. Based on the number of distinct alignment patterns under ML (313 for ITS vs. 1181 for plastid partitions; Table 2), a phylogenetic split *fully* supported by the ITS data but rejected by the plastid data, will, theoretically, be expressed by a $BS_{ML}$ of about 20, the alternative favoured by the plastid data receiving $BS_{ML}$ of 80. The experience with various data sets shows that conflicting signal from one partition is easily overruled by the other partitions in concatenated data sets XXXX, and that a wrong branch inflicted by a single gene can be manifested by adding more data (additional partitions; Delsuc 2005). This can be exemplarily shown for the multigene data set of Li et al. Li 2004 that is the basis for the Fagales phylogeny (Table 5). The relatively low support ($BS_{ML}$ = 67) for the sister relationship between Myricaceae and Betulaceae and allies in the original analysis of Sauquet et al. Sauquet 2012, is due to the fact that this putative clade is only sustained by a weak signal from the matrix. The placement, is, however, stable against deletion of one partition (Fig. 3), but it is also in conflict with the favoured relationship based on the 6-gene data of Li et al. Li 2004, which favours a clade comprising Myricaceae and Juglandaceae. The latter is however mostly due to the signal from the nuclear-encoded 18S rDNA, which is virtually indistinct in Myricaceae and Juglandaceae; and, if not combined with other data, replaces the all-Fagales root one node up (Fig. 4). The relative low supports for the leaves within *Nothofagus* subtree are also due to general weak differentiation, in particular between members of the same subgenus from the same biogeographic region. Five of the concatenated

six partitions lack sufficient signal under ML to resolve relationships in *Nothofagus* outside the subtree of subgenus *Lophozonia* (Fig. 3). The topology of the subtree that includes subgenera *Brassospora, Fuscospora,* and *Nothofagus* relies mostly on signal from the *atp*B-*rbc*L spacer. Further investigation of the BS results of the 5-partition matrix excluding the *atp*B-*rbc*L spacer, shows that although the ML optimisation has difficulties to find an optimal tree, at least the subgenera are fairly supported ($BS_{ML}$ = 57/73/55). The taxa that are the main source of ambiguity are *Nothofagus carrii* (sg. *Brassospora,* New Guinea), *N. baumanniae* (sg. *Brassospora,* New Caledonia), and *N. cliffortioides* (sg. *Fuscospora,* New Zealand); these are also three of the four taxa which highest iDV based on the concatenated matrix and also among the worst sampled: in addition to *atp*B-*rbc*L only data from the *rbc*L gene (c. 100 nt) have been included, which has limited discriminative power for intragenic relationships in *Nothofagus* and other genera of the Fagales.

**TODO Figure 2.** Phylogenetic network based on (Hamming) pairwise distances, with ML bootstrap support annotated for selected phylogenetic splits. [ADD-ON Distance Heat-Map]

**TODO Figure 3.** The six preferred topologies based on data sets with one partition excluded from analysis.

Figure 2 Add-on



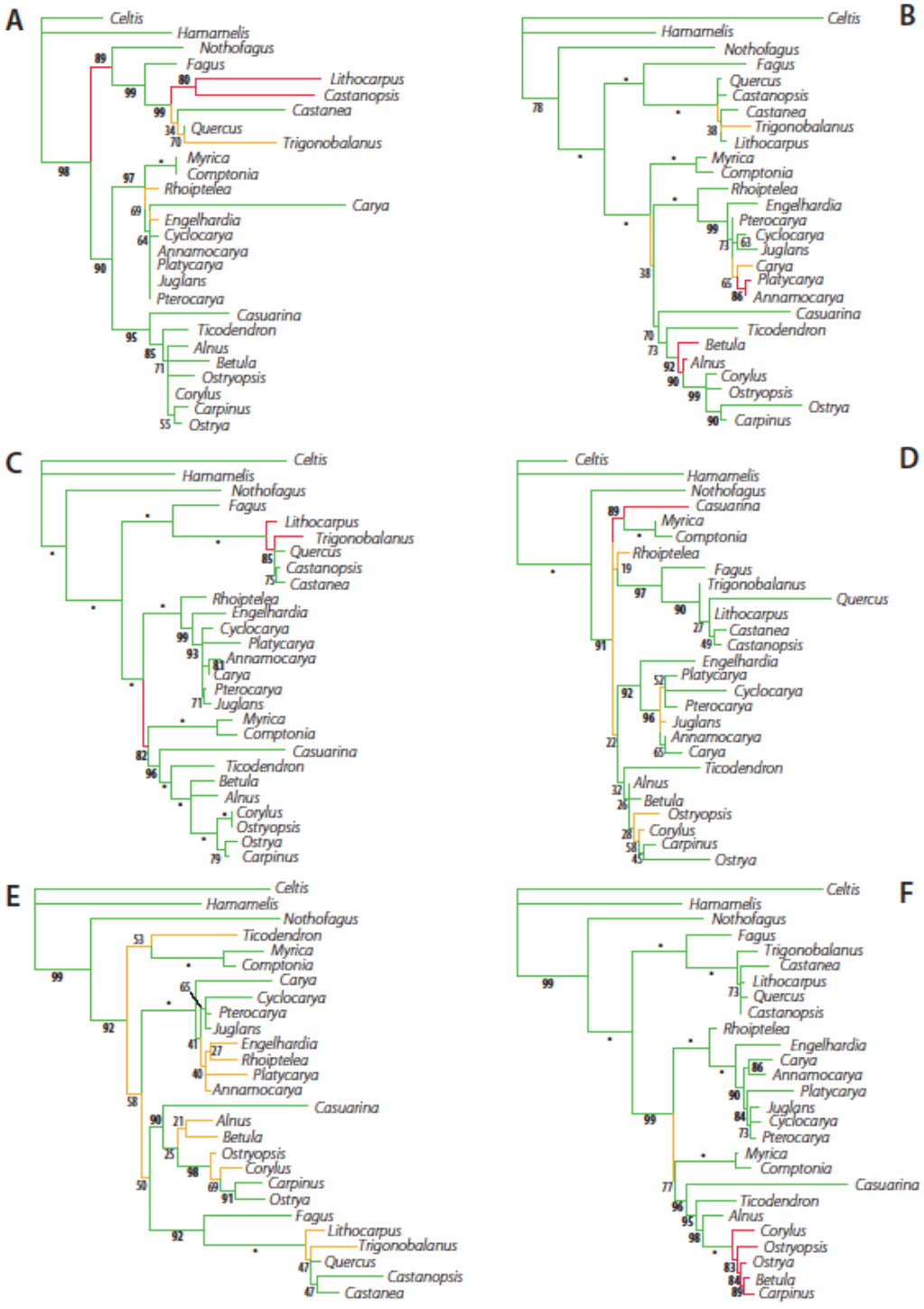



**Figure 4.** Topological phenomena inflicted by 18S rDNA, and other markers, in a genus-level all-Fagales tree (Li 2004 matrix). ML trees based on single-gene matrices; numbers at branches indicate bootstrap support under ML ($BS_{ML}$). $BS_{ML}$ = 100 indicated by asterisks. Green edges, as in 6-gene tree. Orange, incongruent to 6-gene tree but with low support, Red, incongruent with high support **A,** 18S nrDNA data: the outgroup-inferred ingroup root moves, with the result that *Nothofagus* is recognized as sister to the Fagaceae and not all Fagales. **B,** *atp*B data. **C,** *matK* data **D,** *matR* data (mtDNA). **E,** *rbcL* data. **F,** *trnL* data.

## *Intra- and inter-specific, inter-continental and subgeneric differentiation in* Nothofagus

The overall highest intra- and interspecific variation is found within the ITS region of *Nothofagus*. All chloroplast regions covered in the re-assessment show no significant (diagnostic) interspecific variation, when from the same provenance and subgenus. This directly explains the high backbone support in the *Nothofagus* subtree: identical sequences are not easily separated by a phylogenetic tree.

Broader sampling of individuals per species results in marked intra-specific (inter-individual) plastid variation, obscuring inter-specific variation for the most part below the level of subgenera. The real extent of plastid variability in subgenus *Brassospora* has yet to be determined. Only for one plastid region, the *rbcL* gene, and species, *N. discoidea* (New Caledonia), two more than 15-years old accessions Martin Dowd 1993 Setogushi 1997 are available differing at 11 positions; for comparison, the *maximum* difference between a New Guinean, *B. grandis,* and New Caledonian *rbc*L, *B. aequilateralis,* are five substitutions. However, only the patterns at a single position may be significant in either case, the remainder show characteristics of sequencing and editing errors known for old sequences (Grimm 2003) and would need to be verified first by new data (see ES).

Plastid haplotypes are identical in species of sg. *Lophozonia* from South America and New Zealand (*psb*BTNH) and, essentially, of sg. *Brassospora* from New Caledonia and New Guinea (*atp*B-*rbc*L*; rbc*L; Fig. 6*)*. This either means that fixation rates within these genera is extremely low, or that the modern distribution range was shaped only recently (Neogene or later; cf. Svenson 2001). The shared *atp*B-*rbc*L and *rbc*L haplotypes of New Guinean and New Caledonian *Brassospora* explain the high support for the subgenus' root. Overall sequence divergence between *psb*BTNH, *atp*B-*rbc*L/*rbc*L (limited, mostly ≥ 15 years old

data) and *trn*L/*trn*L-*trn*F haplotypes from South America vs. Australia/New Guinea vs. New Zealand/New Caledonia within the subgenera *Lophozonia, Fuscospora,* and *Brassospora* is lower than or equal to intra-specific variation documented for the South Amercian subgenus *Nothofagus,* and the disjunct subgenus *Lophozonia* to some degree, based on more recent data (Fig. 5; Acosta Premoli 2010 Mathiasen Premoli 2009 Knapp 2005). Same holds for the *trn*H-*psb*A spacer (recent data only; with focus on subgenus *Nothofagus*). Two observations regarding the evolutionary unfolding of the genus *Nothofagus* can be straightforwardly done (Fig. 5): (*i*) The split best represented in the collected plastid data is the one between subgenus *Lophozonia* and the remainder of the genus. (*ii*) Within subgenus *Fuscospora,* the South American species *N. alessandri* is generally less distinct to the consensus of its putative sister lineage, subgenus *Nothofagus* (also South America). This general pattern is most pronounced in the *trn*L/*trn*L-*trn*F regions and the *atp*B-*rbc*L spacer. Subgenus-specific sequence patterns are also found for the other subgenera, to a lesser extent, but still predominant in the data. Together with the non-existent differentiation between species of the same subgenus and regions, this results in the high supports for the roots of the subgenera if a matrix combining *atp*B-*rcb*L and *trn*L/*trn*L-*trn*F data is used. Only based on the plastid coding regions (*psb*BTNH, *rbc*L) and the relatively conserved non-coding *trn*L intron, a Group I intron with structural constraints REF, the sequence of *N. alessandri* is closest to the consensus of all *Nothofagus* lineages. The *atp*B-*rbc*L/*rbc*L data promotes a split between *Nothofagus-Brassospora* and *Lophozonia-Fuscospora*, which, together with the strong signal to separate *Lophozonia* from the remainder of the genus, can be only be resolved by the well-known *Nothofagus* subtree that places *Lophozonia* as sister to a clade comprising *Fuscospora* and the sister pair *Nothofagus* and *Brassospora* Manos 1997 Svenson 2001 Knapp 2005 Sauquet 2012. The deep divergence of *Lophozonia* finds further support from duplication patterns, but, similar patterns also indicate the relative distinctness of haplotypes expressed as intra-specific variation in subgenus *Nothofagus.* Furthermore, in the *rbc*L gene only two positions actually

support the *Nothofagus-Brassospora* / *Lophozonia-Fuscospora* split and compete with two positions that supporting a split between *Fuscospora* species from New Zealand and all or only the New Caledonian *Brassospora* species. Most unfortunately, the more informative plastid region providing a more stable and reliable signal (*trn*L/*trn*L-*trn*F) is not known for the latter subgenus. A further potentially informative gene region will be the *psb*BTNH gene complex which exhibits a remarkable polymorphism in subgenus *Nothofagus* Premoli Acosta. If *Brassospora* is the sister lineage of subgenus *Nothofagus*, one could expect based on the available data that their *psb*BTNH region shows a (highly) similar sequence to one of the two major haplotypes in species of subgenus *Nothofagus*.

It is important to note that species from the same subgenus and provenance cannot be distinguished based on their plastid data (Fig. 5; see ES for details). The only region that shows some (see above), but entirely random interspecific divergence is the *atp*B-*rbc*L/*rbc*L, which, for the most part, is based on ≥ 15 years old data that most likely include numerous sequencing and editing artefacts. It is discouraging regarding the reliability of these old sequences that all more recently produced *atp*B-*rbc*L/*rbc*L sequences (Knapp 2005; three *Fuscospora* from New Zealand, two *Lophozonia* each from South America and Australia) show no inter-specific intra-provincial and very limited inter-provincial divergence, as expressed by the young divergence ages obtained in the that study Knapp 2005. Thus, it is obligatory for future analyses that include the *atp*B-*rbc*L/*rbc*L region, or, as in the case of Sauquet et al. Sauquet 2012 rely nearly exclusively on this region (see above; Sauquet 2012, Appendix S1), to verify the old sequences by new data. This will avoid arbitrary long terminal branches based on random PCR-, sequencing and/or editing artefacts that result in instable divergence age estimates.

In contrast to the plastid regions, the ITS region shows a more concise differentiation pattern, down to the species-level in subgenus *Nothofagus*, which is the best sampled Acosta Premoli, and *Lophozonia* (to a lesser degree; Fig. 6). Only *N. betuloides* and *N. dombeyi* of

subgenus *Nothofagus* share identical ITS sequence types. The data basis is sufficient in this subgenus to identify the putative species of two unlabelled accessions from South America (AF480091/AF480092). As in case of the plastids, species grouped in the other two subgenera are highly similar to identical within a certain provenance, but all provenances appear to show diagnostic ITS sequences. It remains, however, to be seen if the signals remain stable if the sampling of subgenera *Fuscospora* and *Brassospora* will be increased. Only for three of the so far sequenced ten species newer data is available for cross-checking the nearly 20 years old, consistently edited original data (one sequence each for a New Guinean *Brassospora* and two New Zealandish *Fuscospora* members, no confirmed data for New Caledonia and the two Australian or South American species of *Fuscospora*). No ITS data is available for those New Caledonian species of subgenus *Brassospora* that are distinct to their New Guinean relatives based on the *atp*B-*rbc*L/*rbc*L region. The substantial editing effort in the original data (Manos 1997, which is also the only data used in Sauquet 2012) is illustrated by the fact that (*i*) potential pseudo-mutations are not as random as in the equally old *atp*B-*rbc*L and *rbc*L data, (*ii*) inter-specific variation is lower in ITS than in *atp*B-*rbc*L/*rbc*L, but still more diagnostic, and (*iii*) the overall divergence is higher in the ITS than in the plastid markers. Nevertheless, by comparison with more recent data, at least five systematic editing artefacts can be found, the most obvious being the 31 nt-long pseudo-deletion in the 5.8S rDNA Manos 1997, not found in any later sequence.

**Table 5.** Phylogenetic splits that received high support from one partition in the matrix of Li et al. (2004), which are in conflict with the preferred tree based on the concatenated, 6-gene data.

| | ***Nothofagus*** | **Fagaceae** | **Myricaceae** | ***Casuarina*** | ***Ticodendron*** | **Betulaceae** | ***Rhoiptelea*** | **Juglandaceae** |
|---|---|---|---|---|---|---|---|---|
| **6-gene matrix** | First Fagales branch | Monophyletic<br><br>*Fagus* sister to other Fagaceae | Monophyletic | Sister to *Ticodendron* + Betulaceae | Sister to Betulaceae | Monophyletic<br><br>Clade comprising *Carpinus, Corylus, Ostrya* and *Osytryopsis* | … sister to Juglandaceae | *Engelhardia* sister to remaining Juglandaceae |
| **18S** | **…sister to Fagaceae!** | *Cp.*+*Lc.* sister to *Ca.*+(*Q.*+*Tb.*) | Unresolved, in a clade with *Rhoipt.* and Juglandaceae | | | | **Unresolved, in a clade with Myricaceaeand Juglandaceae** | Intrafamiliar relationships unresolved*, in a clade with *Rhoiptelea* and Myricaceae |
| **atpB** | | *Qu,Cp* unresolved; *Ca-Lc-Tb* clade | Sister to CTB clade+RJ clade | | | *Betula* + (Alnus + others) | | Basal trichotomy *Pt.*, (*C.* +(*Ac.*+*Pl.*)), Ju.+Cc. |
| **matK** | | *Lithoc.* sister to others | Sister to CTB clade | | | Basal trichotomy: *Alnus, Betula,* others | | Unresolved polytomy *Cc., Ac.*+*C., Pl., Ju.*+*Pt.* |
| **matR** | | *Trigonob.* sister to others | Sister to *Casuarina* | **Sister to Myricaceae!** | | Basal trichotomy: *Alnus, Betula,* others | **… sister to Fagaceae!** | Basal trichotomy *Ju., Ac.*+*C., Cc.*+*Pl.*+*Pt.* |
| **rbcL** | | *Lithoc.* sister to others | Sister to *Ticodendron* | **Within Betulaceae!** | **Sister to Myricaceae!** | *A.*+(*B.*+*Casuar.*) sister to others | **C. sister to *Ju.-Cc.-Pt.* clade + *Rh.-E.-Ac.-Pl.* clade!** | |
| **trnL** | | *Trigonob.* sister to others | Unresolved | | | Basal polytomy | | *Ac.*+*C.* sister to *Pl.*+(*Cc./Pt./Ju.*) |

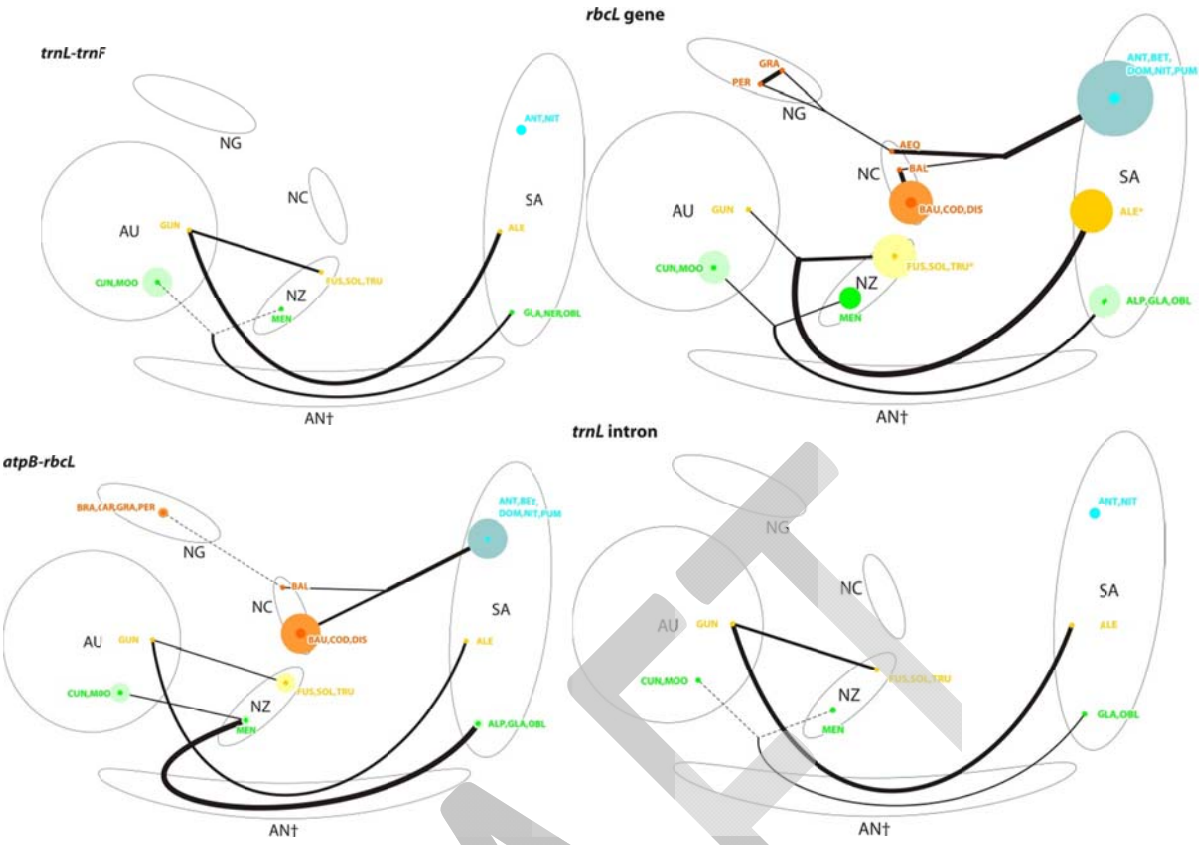



**Figure 5.** Haplotype network-based distances put in a biogeographic framework: plastid data. Line thickness is proportional to the minimum number of reconstructed mutations (1 px = 1 mutational event; stippled lines indicate identical sequences). Circles refer to the intra-provincial variation. Fat coloured, reliable data; light coloured, including potential pseudomutations.

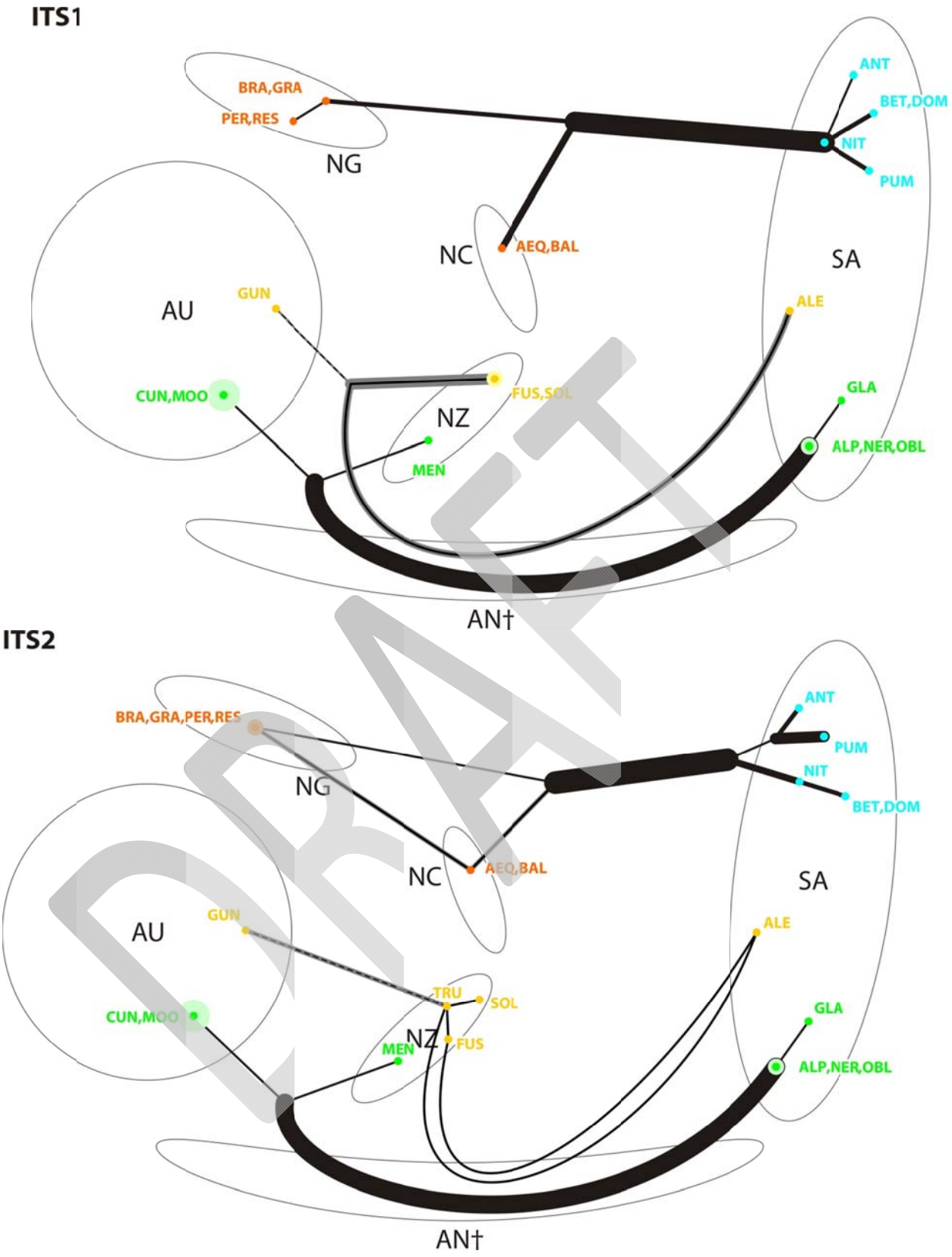



**Figure 6.** Haplotype network-based distances put in a biogeographic framework: ITS data (see Fig. 5).

## *Possible and impossible minimum ages for divergences in* Nothofagus

Figures 5 and 6 demonstrate that all constraints used by Sauquet et al. Sauquet 2012 for the ingroup of *Nothofagus* are hard to reconcile with the actual genetic differentiation patterns in the genus. Primarily the non-coding regions show a significant differentiation between the species today found east of the South Pacific (South America) and those west of the South Pacific (Australia, New Zealand, New Caledonia, New Guinea), divergences that have been constrained with minimum 31.5 Ma ('safe' ingroup fossil constraints) and 40 Ma (geological constraints). Assuming a general low fixation and homogenisation rate in *Nothofagus,* the genetic differentiation found between *Nothofagus* and *Brassospora* and South American and western Pacific members of subgenus *Lophozonia* could well-fit with such an old age. Many *Nothofagus* are ecologically strongly restricted, but equally dominant in their niches than *Fagus* and the temperate Betulaceae and Fagaceae on the northern hemisphere. It can be assumed that in its history *Nothofagus* had episode with extremely large distribution ranges that were closed. The Fagaceae provide numerous examples for low fixation rates. For instance, the disjunct species pair *F. crenata* (Japan) and *F. sylvatica* (western Eurasia) are genetically and morphologically indistinct Denk 2005, but gene flow between both species must have broken down at least at the end of the Pliocene, when the climate in central Asia became too continental for beech trees Denk Grimm 2009. For *Fagus*, the lack of genetic differentiation correlates to phases of wide and closed distribution as evidenced by the fossil record Denk 2004 Denk Grimm 2009. Identical ITS types can also be found in North American and Eurasian white oaks Denk Grimm 2010; the last migration over the North Atlantic Land Bridge stopped about 8 Ma Denk 2011. Gene flow may have been possible via Beringia and Russia/Siberia during warm periods of the Pleistocene, however, this alone cannot explain the lack of genetic differentiation in all white oaks. ADD RESULTS OF NEW BETULACEAE DATING

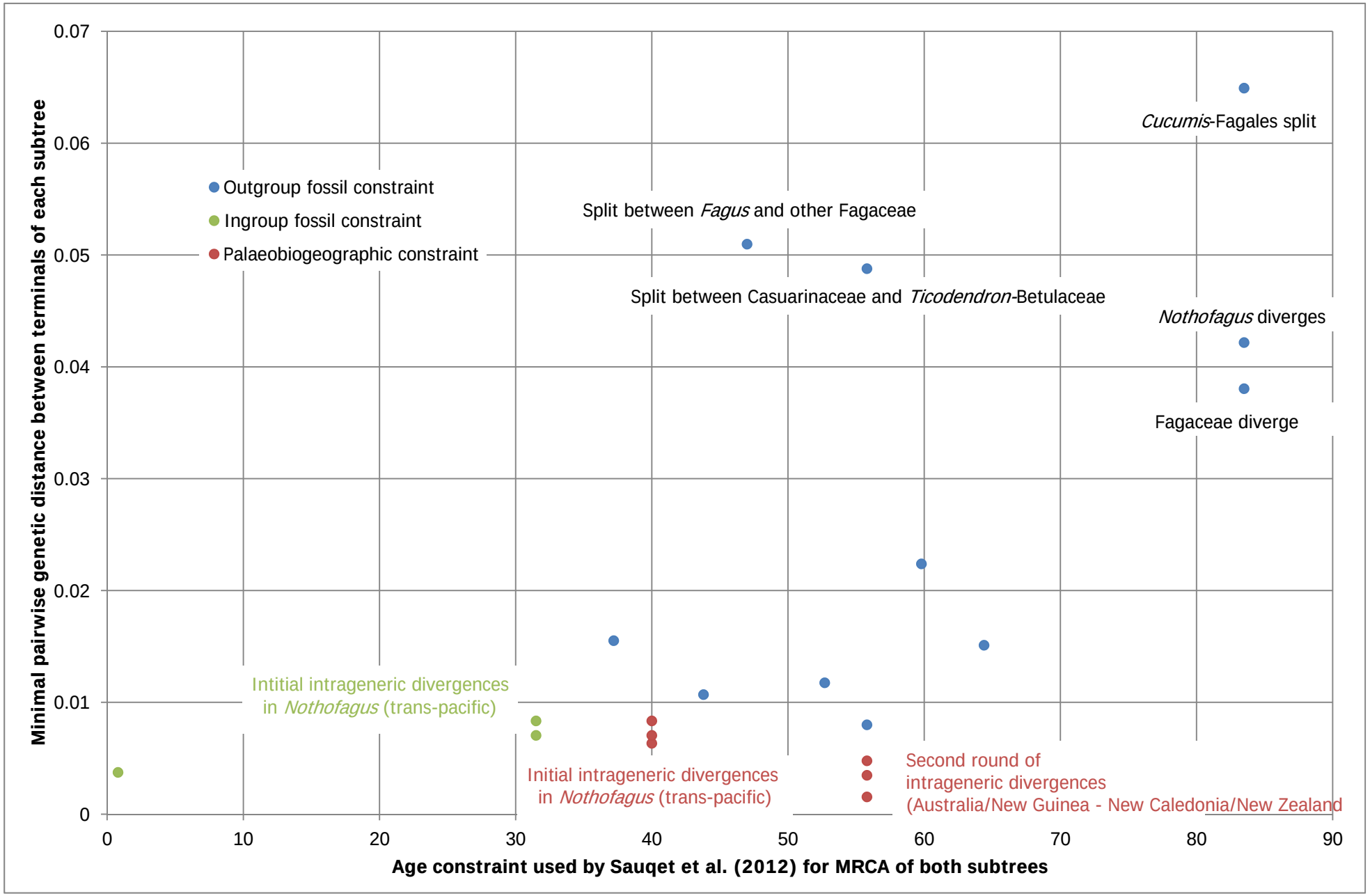
Minimum ages of 55 Ma for divergences between Australia/New Zealand (*Fuscospora, Lophozonia*) and New Caledonia/New Guinea (*Brassospora*) can be straightforwardly rejected based on the fact that genetic differentiation between according species pairs is minimal (Fig. 6), if existent at all (Fig. 5). It is obvious that the sea-ways that formed in the course of the Gondwana break-up were not accompanied by vicariance of species of *Nothofagus* in the newly forming Australasian realm. Instead, one should consider the possibility that New Guinea and New Caledonia were only relatively recently populated by *Nothofagus* and that gene flow persisted between New Zealand and Australia during most part of the Neogene. Fossils of *Brassospora* …

The lower trans-pacific divergences in the ITS and *atp*B-*rbc*L spacer regions of subgenus *Fuscospora* require further investigation about dispersal patterns in this group and its population dynamics compared to other *Nothofagus*. In the generally much more conserved plastid coding regions and *trn*L intron, members of subgenus *Fuscospora* are as a trend more diverged than those of *Lophozonia*. This indicates that fixation rates in this subgenus are not generally lower than in the other subgenera, which would be an alternative explanation for the low relatively low divergence in the non-coding regions. Accordingly, it is not reasonable to assume at all that the basic trans-pacific divergence in this subgenus was co-eval with that in *Lophozonia* or *Brassospora-Nothofagus* (Fig. 7) … ecology of Fuscospora?…

**Figure 7.** Absolute genetic divergence between sister lineages whose minimum ages have been constrained by Sauquet et al. Sauquet 2012

**Link** to commented Appendices S1 and S2 of Sauquet 2012 (data inconsistencies, problems with used fossil constraints):

www.palaeogrimm.org/data/R_origAppS1S2commented.xls


Minimal pairwise genetic distance between terminals of each subtree
0.07
0.06
0.05
0.04
0.03
0.02
0.01
0
0
10
20
30
40
50
60
70
80
90
Age constraint used by Sauqet et al. (2012) for MRCA of both subtrees
Outgroup fossil constraint
Ingroup fossil constraint
Palaeobiogeographic constraint
Cucumis-Fagales split
Split between Fagus and other Fagaceae
Split between Casuarinaceae and Ticodendron-Betulaceae
Nothofagus diverges
Fagaceae diverge
Intitial intrageneric divergences in Nothofagus (trans-pacific)
Initial intrageneric divergences in Nothofagus (trans-pacific)
Second round of intrageneric divergences (Australia/New Guinea - New Caledonia/New Zealand