A peer-reviewed version of this preprint was published in PeerJ on 7 June 2019.

Worth JRP, Liu L, Wei F, Tomaru N. 2019. The complete chloroplast genome of *Fagus crenata* (subgenus *Fagus*) and comparison with *F. engleriana* (subgenus *Engleriana*) PeerJ 7:e7026
https://doi.org/10.7717/peerj.7026

The complete chloroplast genome of *Fagus crenata* (subgenus *Fagus*) and comparison with *F. engleriana* (subgenus *Engleriana*)

James Worth [Corresp., 1], **Luxian Liu** [2], **Nobuhiro Tomaru** [3]

[1] Department of Forest Molecular Genetics and Biotechnology, Forestry and Forest Products Research Institute, Tsukuba, Japan

[2] Key Laboratory of Plant Stress Biology, Laboratory of Plant Germplasm and Genetic Engineering, School of Life Sciences, Henan University, Kaifeng, China

[3] Graduate School of Bioagricultural Sciences, Nagoya University, Nagoya, Japan, Japan

Corresponding Author: James Worth
Email address: jrpw2326@affrc.go.jp

This study reports the whole chloroplast genome of *Fagus crenata* (subgenus *Fagus*), a foundation tree species of Japanese temperate forests. The genome was a total of 158,247 bp in length containing 111 genes. Comparison with the only other published *Fagus* chloroplast genome, *F. engeleriana* (subgenus *Engleriana*) shows that the genomes are relatively conserved with no inversions or rearrangements observed between them and differing by 311 single nucleotide polymorphisms. The six most variable regions between the two genomes were the *psb*K-*psb*I, *trn*G-*psb*fM, *trn*V, *rpl*32, *ndhD*-*psa*C and *ndh*I-*ndh* regions. These highly variable chloroplast regions and the identification of 42 variable chloroplast SSRs found to be shared between the two species will provide useful genetic resources for studies of the inter- and intra-specific genetic structure and diversity of this important northern hemisphere tree genus.

The complete chloroplast genome of *Fagus crenata* (subgenus *Fagus*) and comparison with *F. engleriana* (subgenus *Engleriana*)

Worth J.R.P.[1], Luxian L.[2] and Tomaru, N.[3]

[1] Department of Forest Molecular Genetics and Biotechnology, Forestry and Forest Products Research Institute, Matsunosato 1, Tsukuba, Ibaraki 305-8687, Japan.

[2] Key Laboratory of Plant Stress Biology, Laboratory of Plant Germplasm and Genetic Engineering, School of Life Sciences, Henan University, Kaifeng, 475000, China

[3] Graduate School of Bioagricultural Sciences, Nagoya University, Furo-cho, Chikusa-ku, Nagoya 464-8601, Japan

Corresponding Author: James R.P. Worth

1 Matsunosato, Tsukuba, Ibaraki Prefecture, 305-8687, Japan

Email address: jrpw2326@affrc.go.jp

Abstract

This study reports the whole chloroplast genome of *Fagus crenata* (subgenus *Fagus*), a foundation tree species of Japanese temperate forests. The genome was a total of 158,247 bp in length containing 111 genes. Comparison with the only other published *Fagus* chloroplast genome, *F. engeleriana* (subgenus *Engleriana*) shows that the genomes are relatively conserved with no inversions or rearrangements observed between them and differing by 311 single nucleotide polymorphisms. The six most variable regions between the two genomes were the *psb*K-*psb*I, *trn*G-*psb*fM, *trn*V, *rpl*32, *ndhD*-*psa*C and *ndh*I-*ndh* regions. These highly variable chloroplast regions and the identification of 42 variable chloroplast SSRs found to be shared between the two species will provide useful genetic resources for studies of the inter- and intra-specific genetic structure and diversity of this important northern hemisphere tree genus.

Keywords beech, chloroplast SSRs, Fagaceae phylogeny, *Fagus crenata*, whole chloroplast genome

Introduction

The genus *Fagus* is a major tree of temperate forests of the northern hemisphere with two subgenera recognised, *Engleriana* with three species and *Fagus* with seven species (Oh 2015; Renner et al. 2016). The genus has been the focus of intensive genetic studies over the last 30 years enabling insights into relationships of the extant species (Denk, et al. 2005), the impact of the interglacial-glacial cycles on extant genetic diversity (Fujii *et al.*, 2002; Magri *et al.*, 2006) and predictions of the impacts of ongoing climate change (Csilléry *et al.*, 2014). However, despite the significance of the genus there remains a dearth of Next Generation Sequencing based-genetic resources for *Fagus*, including for the chloroplast genome, with the whole chloroplast genome of only a single species, *F. engleriana* of subgenus *Engleriana* (Yang *et al.*, 2018), so far published.

This study reports the whole chloroplast genome of the Japanese endemic *F. crenata,* the first reported of subgenus *Fagus*. This species is a foundation tree of Japan's cool temperate forest ecosystem and is distributed widely from southern Kyushu (31.4° N 130.8° E) to southern Hokkaido (42.8° N 140.2° E). Phylogeographic studies using Sanger sequencing of small portions of the chloroplast genome have revealed strong geographic structuring of chloroplast haplotypes (Fujii *et al.*, 2002), that combined with fossil pollen data (Tsukada, 1982), suggests that the species persisted in multiple coastal refugia and has occupied most of its current wide geographic range in the postglacial. Here we report the whole chloroplast genome sequence of *F. crenata* and compare it to the genome of *F. engleriana* (subgenus *Engleriana*). This data will be a useful genetic resource for investigating the phylogenetic relationship of *Fagus* and for developing chloroplast genetic markers, including both single nucleotide polymorphisms and SSR markers.

Materials and Methods

Whole genomic DNA was extracted from a single sample of *Fagus crenata* collected from Daisengen Peak, Hokkaido, Japan (41.616° N - 140.1333° E) representing the *F. crenata* chloroplast haplotype A (following Fujii *et al.* 2002) using a modified CTAB protocol (Doyle, 1990). DNA concentration and quality were assessed by agarose gel electrophoresis and a Qubit 2.0 fluorometer (Life Technologies). A total of 9 µg of DNA was sent to the Beijing Genomic Institute where short-size Truseq DNA libraries were constructed and paired-end sequencing

(2x100 bp) was performed on an Illumina HiSeq2000 Genome Analyser resulting in a total of 7,223,910 reads. Assembly of chloroplast DNA from the whole genomic sequencing data was undertaken in Novoplasty 2.6.3 (Dierckxsens, Mardulyn, & Smits, 2016), a seed- and-extend algorithm that is designed for the specific purpose of assembling chloroplast genomes from whole genome sequencing data, starting from a chloroplast seed sequence (*trnK-matK* of haplotype A: Genbank accession AB046492). This resulted in nine chloroplast contigs varying in length from 2748 to 43982 bp constructed from 230,360 chloroplast reads (3.19% of the total reads) with an average read coverage of the chloroplast genome of 145. The nine contigs were ordered and oriented using the *Fagus engleriana* whole chloroplast genome (KX852398) as a reference and the complete chloroplast sequence of *F. crenata* was constructed by connecting overlapping terminal sequences. Sanger sequencing of *F. crenata* was undertaken to check the accuracy of assembly of the nine contigs, the joins of the inverted repeat and single copy regions and also the sequences of the most diverged sites between *F. crenata* and *F. engleriana* (see Results and Discussion). A total of 8146 bp was sequenced using fifteen primer pairs and no differences were observed with the *F. crenata* genome apart from those due to inaccurate sequence at the terminal ends of the Sanger sequences.

The annotation of the cp genome was performed using the online program Dual Organellar Genome Annotator (Wyman et al., 2004). Initial annotation, putative starts, stops, and intron positions were determined according to comparisons with homologous genes of *F. engleriana* cp genome using Geneious v9.0.5 (Biomatters, Auckland, New Zealand). The circular gene maps were drawn by the OrganellaGenomeDRAW tool (OGDRAW) following by manual modification (Lohse et al., 2013).

A neighbor joining tree was constructed in Geneious v9.0.5 using the Geneious tree builder algorithm under default parameters from an alignment constructed using MAFFT v7.308 (Katoh *et al.*, 2002) of *Fagus crenata*, *F. engleriana* and representative whole chloroplast genomes of the Fagaceae family and outgroups from Betulaceae and Juglandaceae obtained from Genbank. Chloroplast microsatellite regions shared in both *F. crenata* and *F. engleriana* were searched for using Phobos Tandem Repeat Finder (Mayer, 2008) implemented in Geneious v9.0.5 with a repeat unit length of 1-3 bp and a minimum length of 10 bp. The coding genes, non-coding regions and intron regions were compared between the alignment of the two *Fagus* chloroplast genomes to

detect divergence hotspots. We examined 101 regions (39 coding genes, 52 intergenic spacers, and 10 intron regions) from the two *Fagus* species for nucleotide variability (Pi) values calculated with the DnaSP v5.0 software.

Results and Discussion

The assembled whole chloroplast genome of *Fagus crenata* was a total of 158,247 bp in length (Figure 1: Genbank accession number MH171101) and consisted of an 87,577 bp large single copy region, a 18,928 bp small single copy region and two inverted repeats 25,871 bp in length. The genome contained 111 genes, including 76 protein-coding genes, 31 tRNA genes, and 4 ribosomal RNA genes. The neighbor joining tree showed that *F. crenata* and *F. engleriana* were sisters and formed a clade strongly diverged from a clade containing all other Fagaceae (Figure 2) consistent with previous studies showing the large divergence of *Fagus* from all other Fagaceae genera (Heenan and Smissen 2013). The two *Fagus* chloroplast genomes were relatively conserved (Figure 3) with the IR region more conserved than the LSC and SSC regions. We did not detect either inversions or translocations among the two genome sequences, and no rearrangement occurred in gene organization after verification (Figure 4). The two species differed by 311 single nucleotide polymorphisms or at 0.197% of all aligned 158,106 non-gapped base positions. The nucleotide diversity values between the 101 regions of the two *Fagus* species ranged from 0.0003 (*ycf*2 gene) to 0.0781(*ndh*D-*psa*C) (Figure 5). The six most variable regions were *psb*K-*psb*I, *trn*G-*psb*fM, *trn*V, *rpl*32, *ndhD-psa*C and *ndh*I-*ndh* of which four are located in the LSC region and two are in the SSC region (Figure 5). The highest nucleotide diversities observed between *F. crenata* and *F. engleriana* were higher than observed within other Fagaceae genera including East Asian (Yan et al. 2018) and Mediterranean oaks (Vitelli et al. 2017) consistent with a deep divergence between the chloroplast genomes of the two *Fagus* subgenera. Of a total of 105 chloroplast SSRs identified in the two species, 104 were present in both *F. crenata* and *F. engleriana* and 42 of these displayed size variation between them (Supplementary Table 1). The majority of variable chloroplast SSRs were mono-nucleotide repeats with 61% showing size variation between the two species while only 21% of di-nucleotide repeats and zero of tri-nucleotude repeats did (Figure 6). The length of variable versus non-variable chloroplast SSRs was similar but with a greater length variation for variable SSRs in both *F. crenata* and *F. engleriana* (Figure 7).

Conclusion

Overall, the chloroplast genome described in this study will provide a useful genetic resource for future studies into the inter- and intra-specific genetic structure and diversity of the foundation temperate tree genus *Fagus*.

Acknowledgements

We would like to thank fellow lab members for their advice on this study.

Funding

This work was supported by the Japanese Society for the Promotion of Science Grant-in-Aid for Young Scientists A (Grant number 16748931); and a Forestry and Forest Products Research Institute grant (Gant number 201430).

Competing interestes

The authors declare no potential conflict of interest.

References

Csilléry K, Lalagüe H, Vendramin GG, *et al.* 2014. Detecting short spatial scale local adaptation and epistatic selection in climate-related candidate genes in European beech (*Fagus sylvatica*) populations. *Molecular Ecology* 23: 4696–4708.

Darling AC, Mau B, Blattner FR and Perna NT (2004) Mauve: multiple alignment of conserved genomic sequence with rearrangements. *Genome Research* 14: 1394-1403

Denk T, Grimm GW & Hemleben V. 2005. Patterns of molecular and morphological differentiation in *Fagus* (Fagaceae): phylogenetic implications. *American Journal of Botany* 92: 1006–1016.

Dierckxsens N, Mardulyn P & Smits G. 2016. NOVOPlasty: *de novo* assembly of organelle genomes from whole genome data. *Nucleic acids research* 45: e18–e18.

Doyle JJ. 1990. Isolation of plant DNA from fresh tissue. *Focus* 12: 13–15.

Fujii N, Tomaru N, Okuyama K, *et al.* 2002. Chloroplast DNA phylogeography of *Fagus crenata* (Fagaceae) in Japan. *Plant Systematics and Evolution* 232: 21–33.

Heenan PB and Smissen RD (2013) Revised circumscription of *Nothofagus* and recognition of the segregate genera *Fuscospora*, *Lophozonia*, and *Trisyngyne* (Nothofagaceae). *Phytotaxa* 146: 1-31.

Katoh K, Misawa K, Kuma K, *et al.* 2002. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic acids research* 30: 3059–3066.

Lohse, M., Drechsel, O., & Bock, R. (2007). OrganellarGenomeDRAW (OGDRAW): a tool for the easy generation of high-quality custom graphical maps of plastid and mitochondrial genomes. *Current genetics*, *52*(5-6), 267-274.

Magri D, Vendramin GG, Comps B, *et al.* 2006. A new scenario for the Quaternary history of European beech populations: palaeobotanical evidence and genetic consequences, A new scenario for the Quaternary history of European beech populations: palaeobotanical evidence and genetic consequences. *New Phytologist*, 171: 199–221

Mayer C. 2008. Phobos, a tandem repeat search tool for complete genomes. *Version* 3: 12.

Oh SH. 2015. Sea, wind, or bird: Origin of *Fagus multinervis* (Fagaceae) inferred from chloroplast DNA sequences. *Korean Journal of Plant Taxonomy* 45: 213–220.

Renner SS, Grimm GW, Kapli P, *et al.* 2016. Species relationships and divergence times in beeches: new insights from the inclusion of 53 young and old fossils in a birth–death clock model. *Phil. Trans. R. Soc. B* 371: 20150135.

Tsukada M. 1982. Late-Quaternary development of the *Fagus* forest in the Japanese Archipelago. *Japanese Journal of Ecology* 32: 113–118.

Vitelli M, Vessella F, Cardoni S, Pollegioni P, Denk T, Grimm GW and Simeone MC (2017) Phylogeographic structuring of plastome diversity in Mediteranean oaks (*Quercus* Group Ilex, Fagaceae). *Tree Genetics and Genomes* 13: 3

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

193 Yan M, Yanshi X, Liu R, Deng M and Song J (2018) The application and limitation of universal
194 chloroplast markers in discriminating East Asian evergreen oaks. *Frontiers in Plant Science* 9:
195 569.

196 Yang Y, Zhu J, Feng L, *et al.* 2018. Plastid genome comparative and phylogenetic analyses of
197 the key genera in Fagaceae: Highlighting the effect of codon composition bias in phylogenetic
198 inference. *Frontiers in plant science* 9: 82.

199

200

201

202

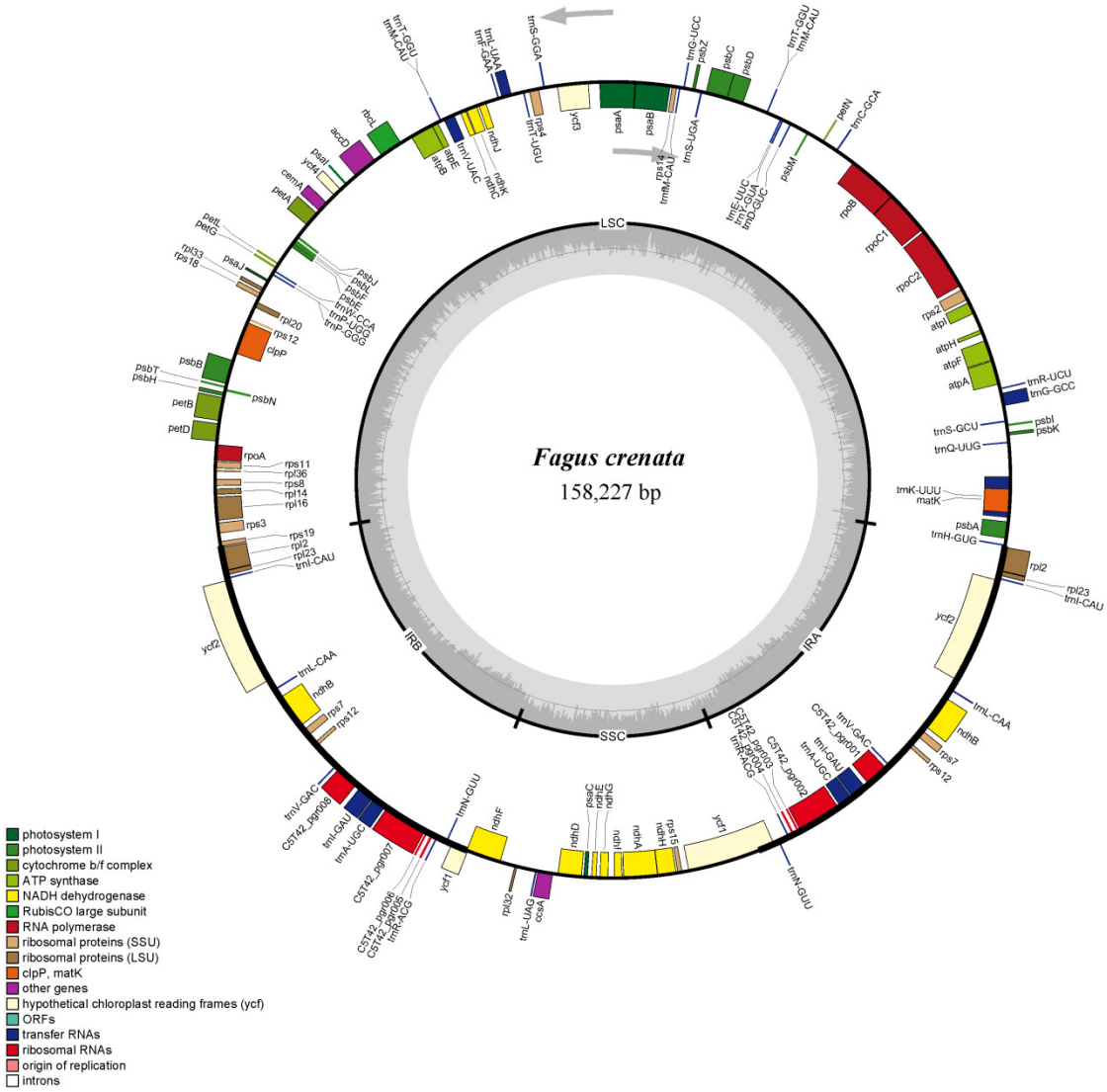


Figure 1. Chloroplast genome maps of *Fagus crenata*. Genes inside the circle are transcribed clockwise, genes outside are transcribed counter-clockwise. The light gray inner circle corresponds to the AT content, the dark gray to the GC content. Genes belonging to different functional groups are shown in different colors.

Figure 2. Neighbor joining phylogenetic tree of *Fagus crenata*, *F. engleriana* and representative genera of the Fagaceae family and outgroups from Betulaceae and Juglandaceae. The Genbank accession number of each chloroplast genome is shown after the species name.

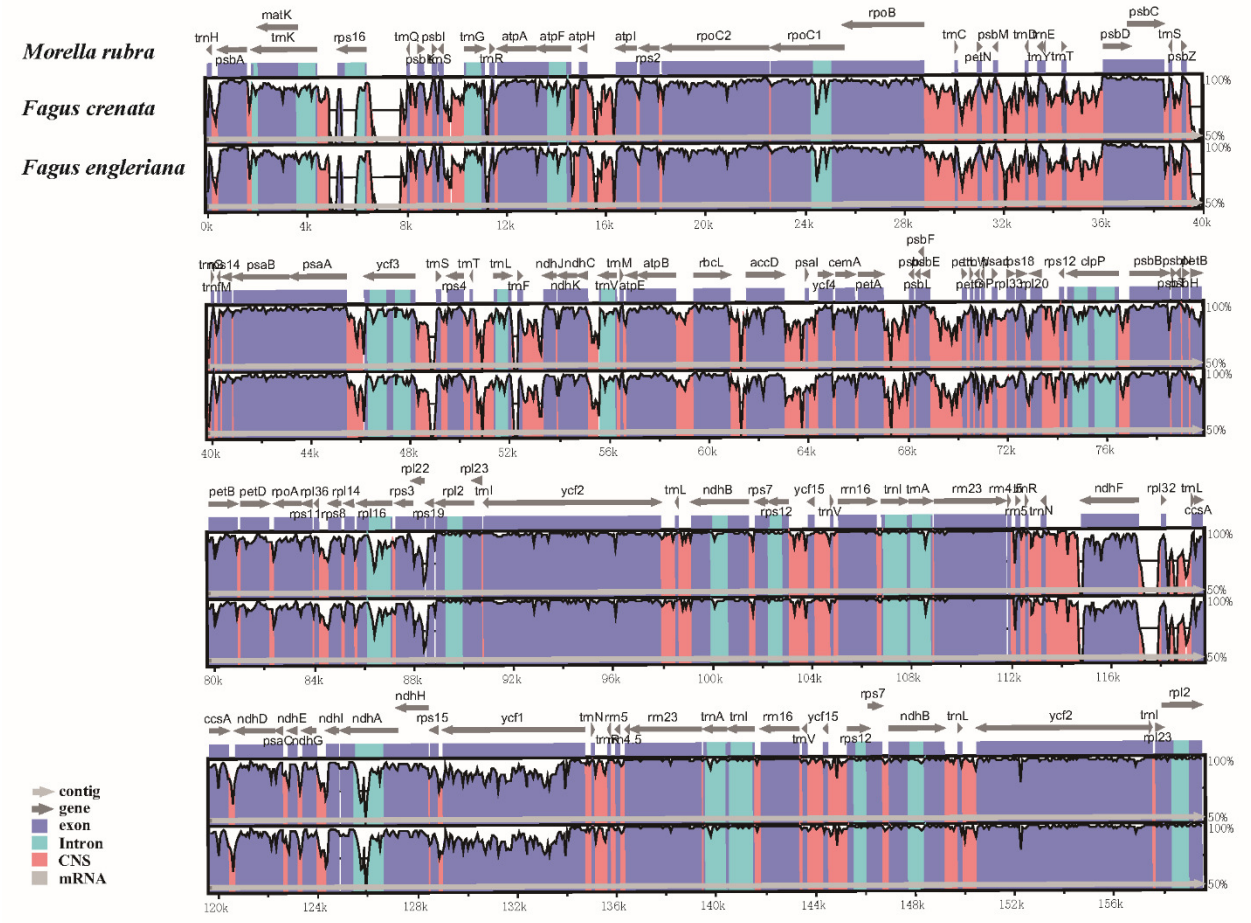


Figure 3. Visualization of alignment of the two *Fagus* chloroplast genome sequences, with *Morella rubra* (Myricaceae, Fagales) as a reference. The horizontal axis indicates the coordinates within the chloroplast genome. The vertical scale indicates the percentage of identity, ranging from 50 to 100 %. Genome regions are color coded as protein coding, intron, mRNA, and conserved non-coding sequence (CNS).

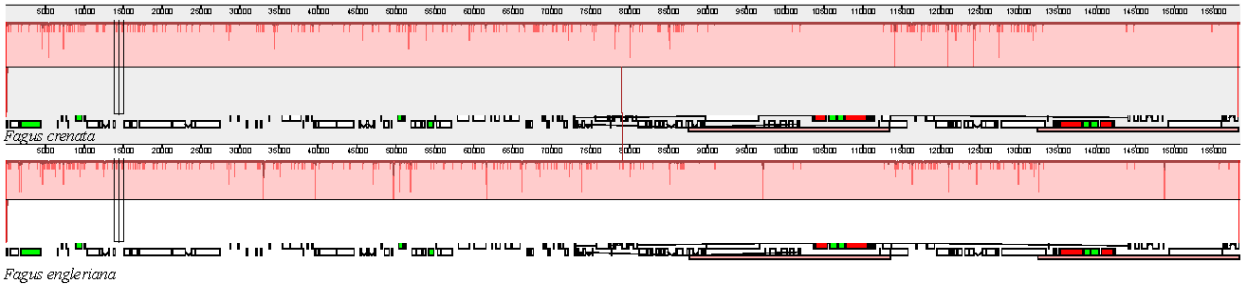


Figure 4 A MAUVE (Darling et al. 2004) alignment of *Fagus crenata* and *F. engleriana* chloroplast genomes showing the lack of re-arrangements between the chloroplast genomes of the two species. The *Fagus crenata* genome is shown at top as the reference. Within each of the alignment, local collinear blocks are represented by blocks of the same color connected by lines.

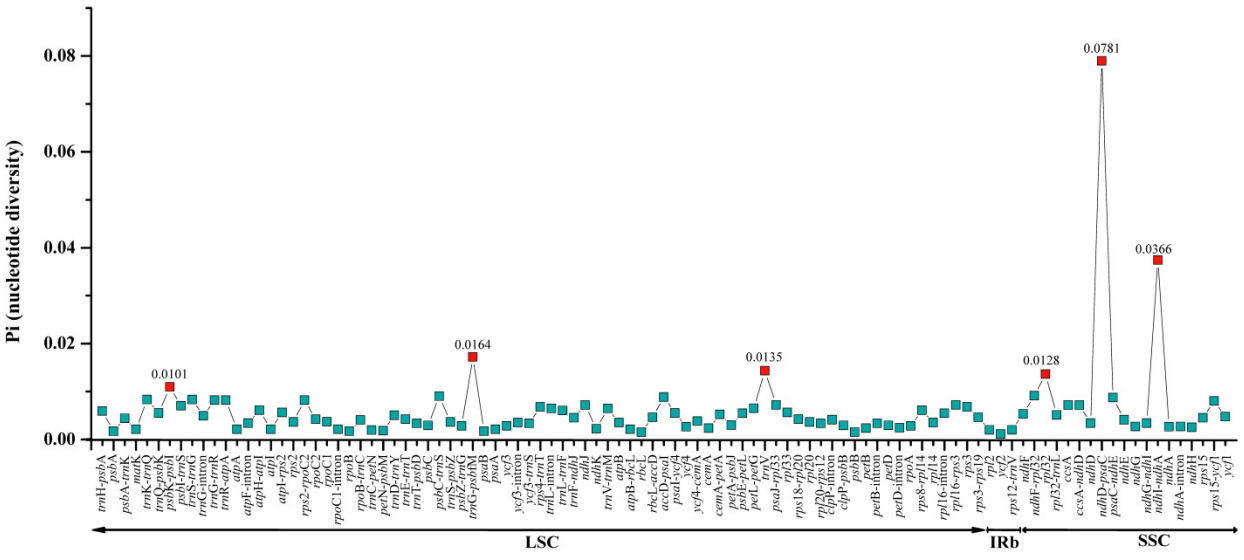


Figure 5. Comparative analysis of the nucleotide diversity (Pi) values between the two *Fagus* species.

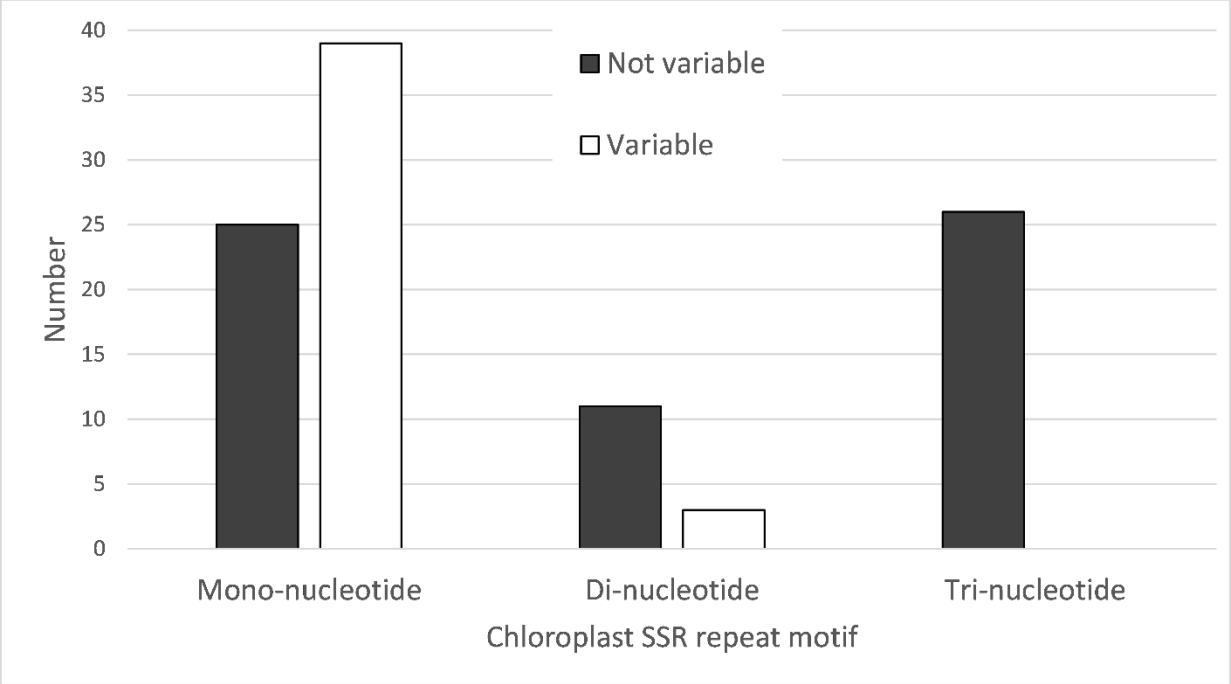


Figure 6. The number of mono-, di-, and tri-nucleotide repeats of the total 104 chloroplast microsatellites over 10 bp in length shared in *Fagus crenata* and *F. engleriana*. The number of these chloroplast microsatellites that displayed no size variation between the two species is indicated by the black bars while those that did are indicated by white bars. Note that the number of variable tri-nucleotide chloroplast SSRs was zero.

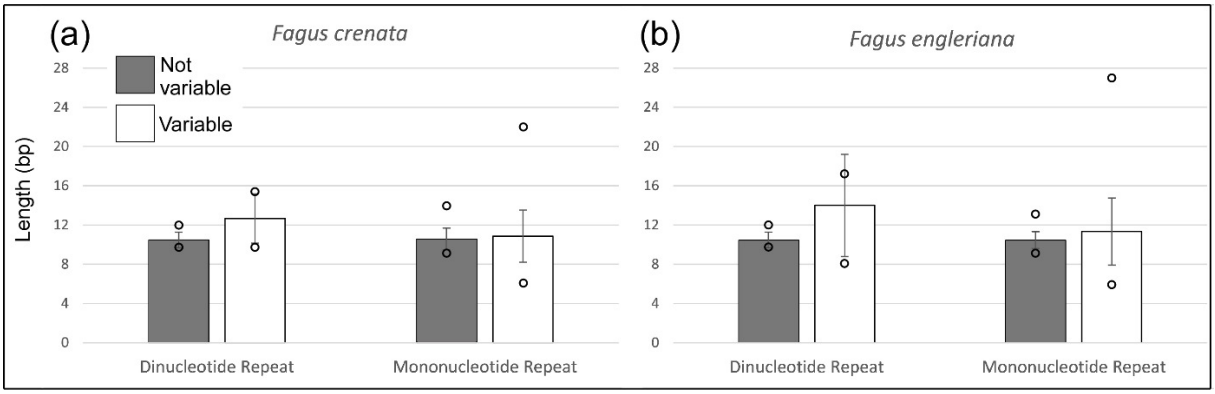
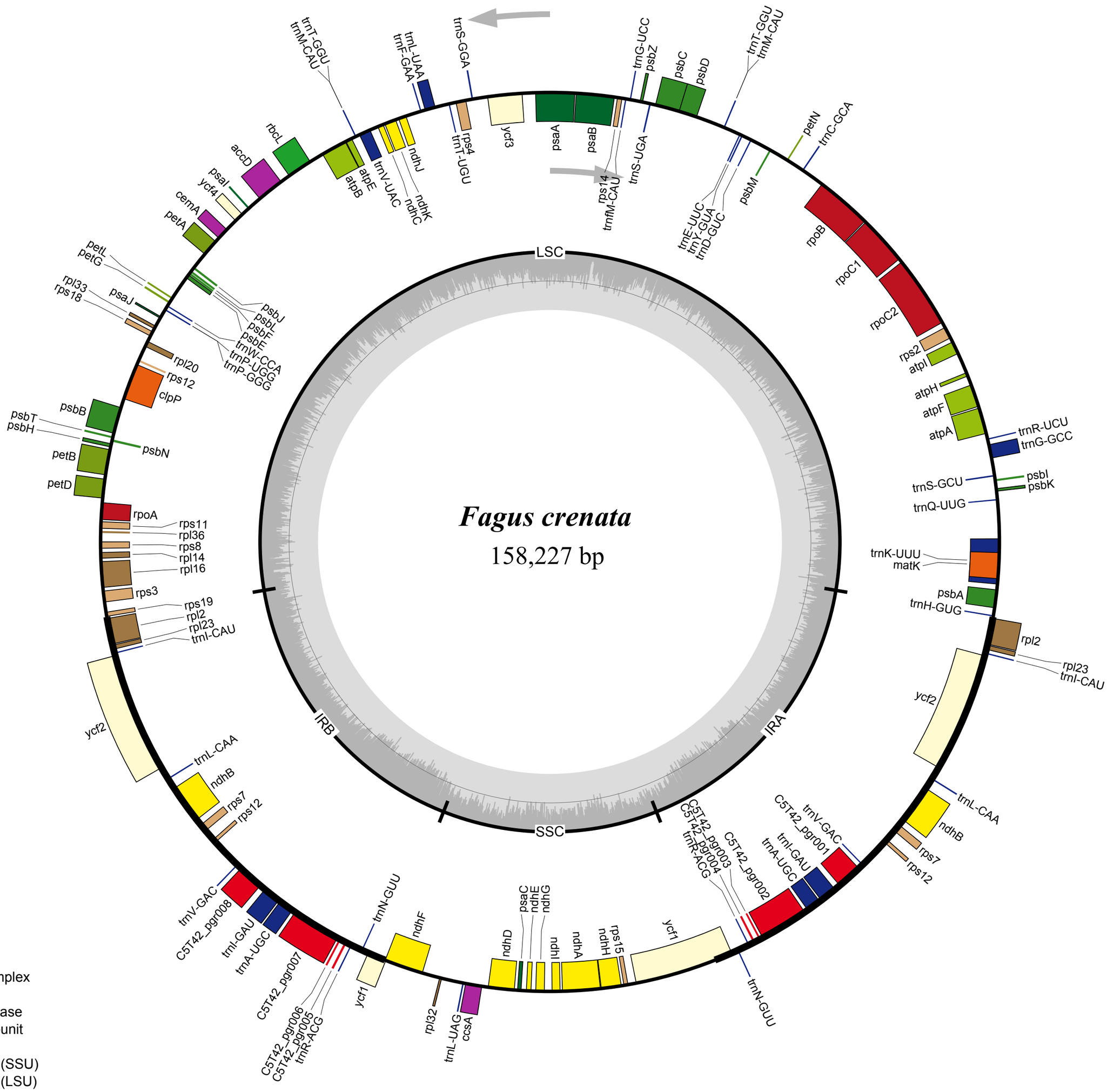


Figure 7 The average length (bp) of variable versus non-variable chloroplast SSRs for both mono- and di- nucleotide repeat motif types observed in both (a) *Fagus crenata* and (b) *F. engleriana* including the standard deviation (error bars) and minimum and maximum lengths (empty circles).

Figure 1(on next page)

Chloroplast genome map of *Fagus crenata*



- photosystem I
- photosystem II
- cytochrome b/f complex
- ATP synthase
- NADH dehydrogenase
- RubisCO large subunit
- RNA polymerase
- ribosomal proteins (SSU)
- ribosomal proteins (LSU)
- clpP, matK
- other genes
- hypothetical chloroplast reading frames (ycf)
- ORFs
- transfer RNAs
- ribosomal RNAs
- origin of replication
- introns

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.27473v1 | CC BY 4.0 Open Access | rec: 10 Jan 2019, publ: 10 Jan 2019

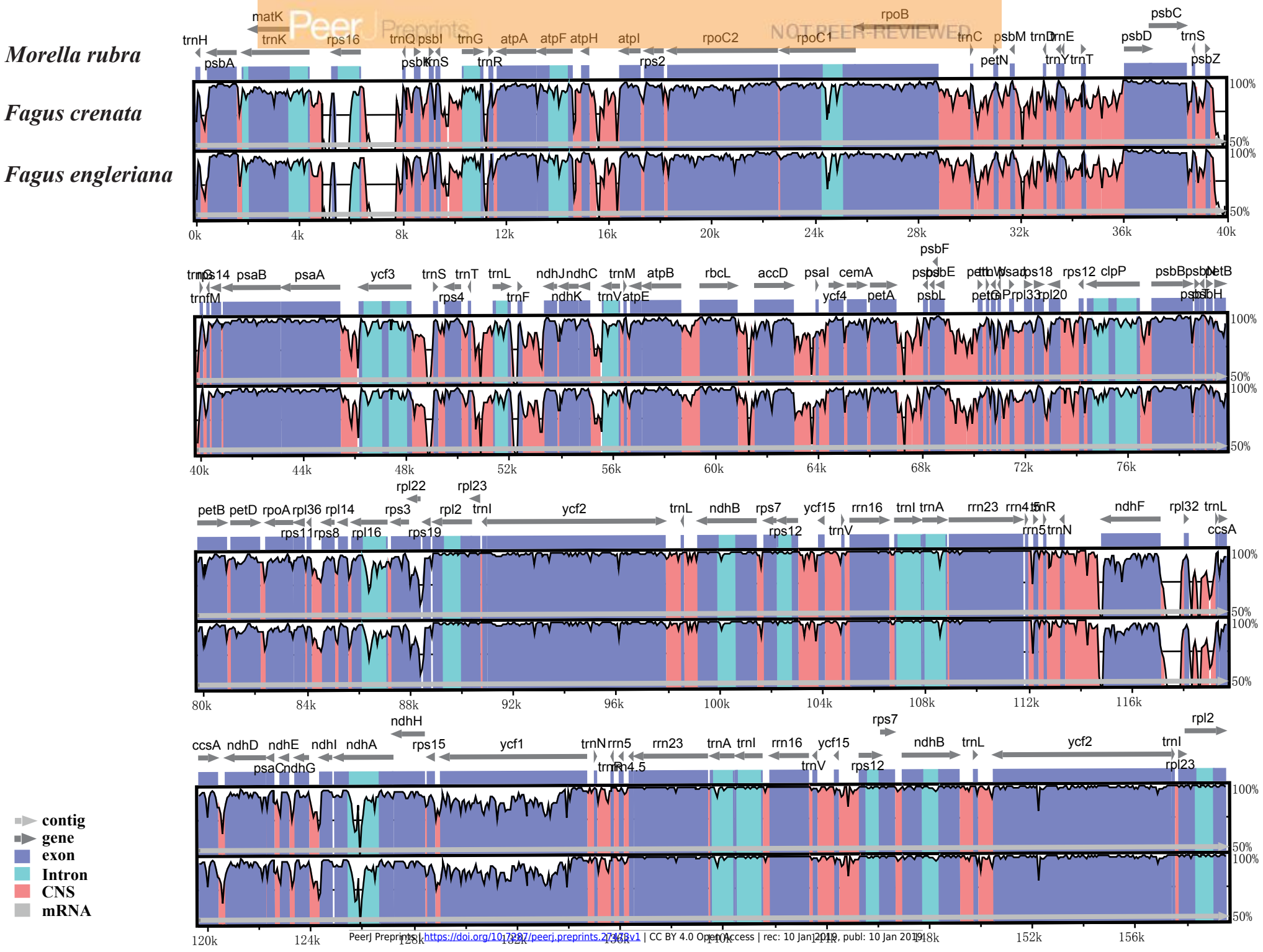
Figure 2(on next page)

Neighbor joining phylogenetic tree of *Fagus crenata*, *F. engleriana* and representative genera of the Fagaceae family and outgroups from Betulaceae and Juglandaceae

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.27473v1 | CC BY 4.0 Open Access | rec: 10 Jan 2019, publ: 10 Jan 2019

Figure 3(on next page)

Visualization of alignment of the two *Fagus* chloroplast genome sequences, with *Morella rubra* (Myricaceae, Fagales) as a reference



PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.27478v1 | CC BY 4.0 Open Access | rec: 10 Jan 2019, publ: 10 Jan 2019

Figure 4

A MAUVE (Darling et al. 2004) alignment of *Fagus crenata* and *F. engleriana* chloroplast genomes showing the lack of re-arrangements between the chloroplast genomes of the two species.

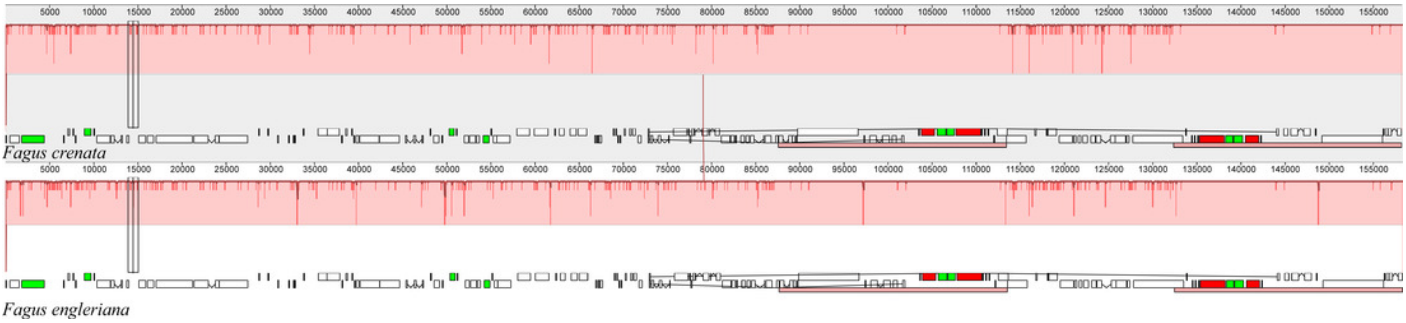
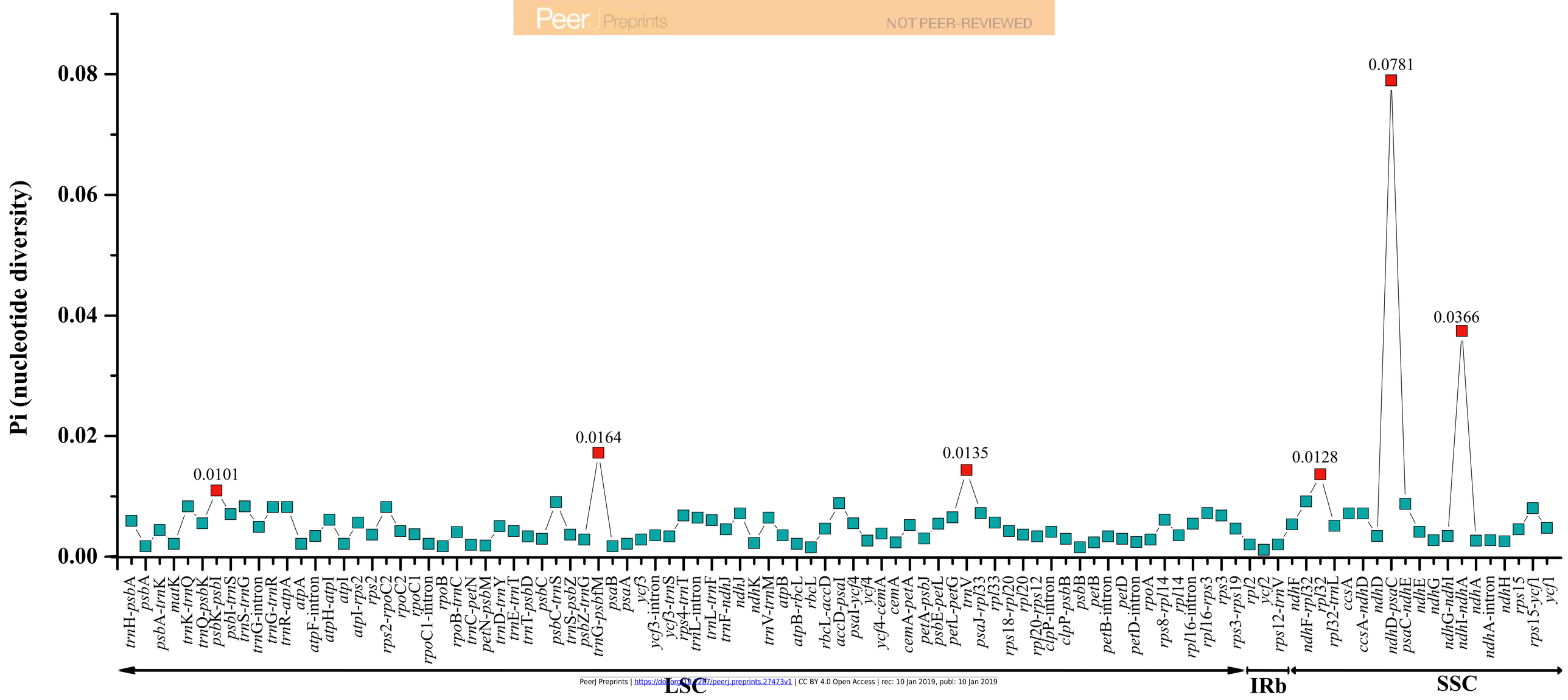


Figure 5(on next page)

Comparative analysis of the nucleotide diversity (Pi) values between the two *Fagus* species



PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.27473v1 | CC BY 4.0 Open Access | rec: 10 Jan 2019, publ: 10 Jan 2019

Figure 6

The number of mono-, di-, and tri-nucleotide repeats of the total 104 chloroplast microsatellites over 10 bp in length shared in *Fagus crenata* and *F. engleriana*

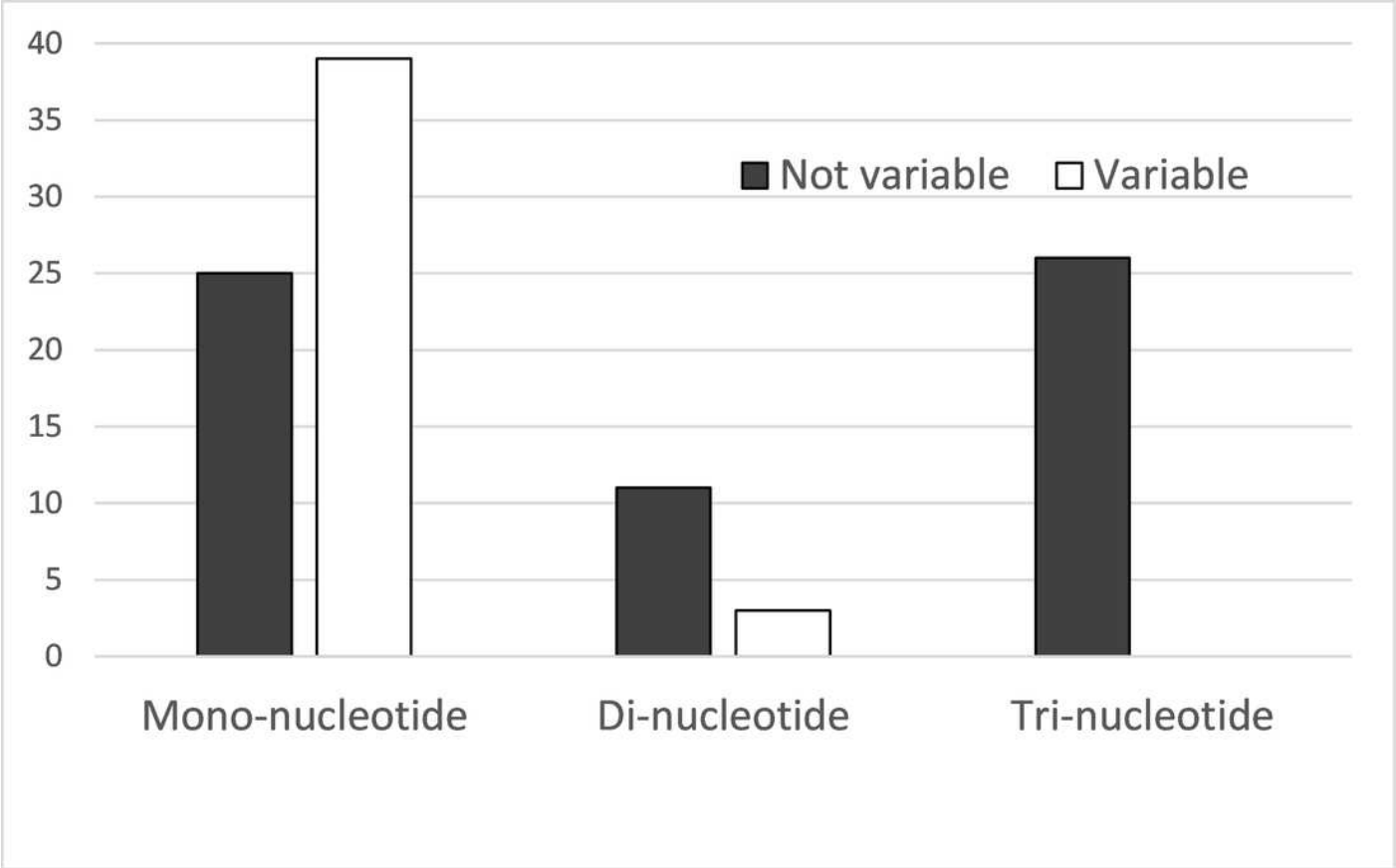


Figure 7

The average length (bp) of variable versus non-variable chloroplast SSRs for both mono- and di- nucleotide repeat motif types observed in both (a) *Fagus crenata* and (b) *F. engleriana* including the standard deviation (error bars) and minimum an