Department of **Biodiversity, Conservation and Attractions**

Biodiversity and Conservation Science

RESEARCH REPORT

Genetic assessment of the Busselton populations of *Conospermum caeruleum*

Donna Bradbury, Rachel Binks and Margaret Byrne

Biodiversity and Conservation Science

Department of Biodiversity, Conservation and Attractions

Keiran McNamara Conservation Science Centre, 17 Dick Perry Avenue, Kensington, WA, 6151

Email: Donna.Bradbury@dbca.wa.gov.au

16th December 2019

A report for Water Corporation and Department of Main Roads

EXECUTIVE SUMMARY

Background and aims

- *Conospermum caeruleum* subspecies '*Busselton*' is a recently discovered, proposed taxon within the *C. caeruleum* species complex, growing in coastal limestone-associated sands, primarily within the City of Busselton. Most known populations are subject to impending disturbance activity and/or are without formal conservation protection.
- If subsp*. 'Busselton'* is a distinct genetic entity it would be considered rare and threatened, requiring urgent conservation genetic management activities to offset disturbance to a significant portion of its natural range. If subsp. *'Busselton'* is not genetically distinct, it would be considered widespread and not under imminent threat from development within the City of Busselton.
- Existing taxonomy within *C. caeruleum* is ambiguous due to highly variable leaf morphology and growth form, requiring genetic data to clarify relationships.
- This research aims to clarify genetic relationships within *C. caeruleum*, and specifically of subsp. '*Busselton*', to determine the most appropriate conservation genetic management actions.

Methods

- Diversity Arrays Technology sequencing (DArTseq) was used to assess genome-wide Single Nucleotide Polymorphism (SNP) variation within *C. caeruleum,* based on 229 individuals from 1-5 populations of each currently-described subspecies (subsp. *caeruleum*, subsp. *oblanceolatum*, subsp. *marginatum*, subsp. *debile*, subsp. *spathulatum*, subsp. '*Busselton*', subsp. '*Whicher*').
- Phylogenetic (maximum likelihood (ML)) trees and population genetic analyses (STRUCTURE, principal coordinates analysis (PCoA), F_{ST} estimates) were conducted to visualise and quantify genetic relationships within C*. caeruleum*.

Results

- Overall genetic differentiation and structure (F_{ST}) in the dataset was very high, as was the fixation index (F_{IS}) indicating overall high levels of homozygosity, likely due to high levels of inbreeding.
- The ML tree formed four major phylogenetic clades, corresponding strongly to geographic distribution: (1) subsp. *spathulatum*; (2) subsp. *caeruleum* + subsp. *oblanceolatum*; (3) the southern portion of the southwest region near Scott River ('SW south'), and (4) the northern portion of the southwest region from Busselton to the Margaret River Headwaters ('SW north'), including subsp. '*Busselton*', which was a distinct but nested clade within the 'SW north' clade.

- STRUCTURE analysis identified the same four clusters, where subsp. '*Busselton*' belonged to the same cluster as 'SW north' individuals, but showed almost 100% membership to this cluster, unlike the 'SW north' individuals that were admixed, indicating a different signal of ancestry.
- The PCoA analysis of the whole dataset revealed three major genetic clusters on the first two axes, corresponding to the three major geographic regions: (1) Forest (subsp. *spathulatum*); (2) Albany & Stirlings (subsp. *caeruleum* + subsp. *oblanceolatum*); and (3) the southwest (including all subsp. *debile*, subsp. *marginatum*, subsp. '*Whicher'* and subsp. '*Busselton'*).
- A subsetted PCoA analysis specifically of the focal southwest region revealed three major genetic clusters, corresponding to: (1) 'SW south'; (2) 'SW north'; and (3) subsp. '*Busselton*'.

Conclusions

- Genetic differentiation at the larger scale, and particularly within the southwest, was largely explained by geography rather than by the current taxonomy.
- We consider the dataset to represent the presence of at least three closely-related but distinct *Conospermum* species, corresponding to the three broad geographic regions.
- Due to strong genetic differentiation, morphological separation and ecogeographic patterns, we suggest elevation of subsp. *spathulatum* to species level. For the same reasons, we also suggest that subsp. *oblanceolatum* and subsp. *caeruleum* are collapsed into a single entity, which itself is elevated to species level.
- Based on the genetic analysis presented here, our conservative conclusion is that a third, single, highly morphologically variable *Conospermum* species occupies the southwest between Busselton and Scott River and is comprised of three distinct Management Units (MUs) for conservation. These units are: (1) all populations currently recognised as subsp. '*Busselton*'; (2) all other 'SW north' populations, and (3) all 'SW south' populations.
- The distinction of the Busselton MU also has morphological support, but the other MUs do not.
- With the current information, we do not recommend recognition of the Busselton populations as a distinct subspecies, but it should be managed separately and considered an independent conservation unit until more information is known.
- We found no genetic evidence to support the delineation of subsp. *debile* and subsp. *marginatum* as distinct taxa. We also found no genetic evidence to support the delineation of the population on Sues Road (currently referred to as subsp. '*Whicher*') as a distinct taxon.
- Taxonomic revision of this group, including a thorough morphological analysis, is strongly recommended, and may reveal further characters to define the 'SW north' and 'SW south' MUs, which would warrant elevation of all three units to subspecies level.

- The morphological variation in this group is highly variable and sometimes reflects genetic variation, and sometimes does not.
- Further research is required to better understand the genetic patterns and evolution of this complex group, including that of the Busselton populations.

1. BACKGROUND

Conospermum caeruleum R.Br. (Blue Brother) (Proteaceae) is a small, blue-flowered shrub that occurs in forest and woodland of southwest WA in widespread but scattered populations (Figs 1, 2), some of which are subject to significant impending disturbance activity. It is a morphologically-variable species complex, where one extinct (subsp. *contortum*) and five extant subspecies (subsp. *caeruleum*, subsp. *oblanceolatum*, subsp. *spathulatum*, subsp. *marginatum*, subsp. *debile*) were recognised in the most recent formal taxonomic revision (Bennett 1995). The subspecies are distinguished by leaf morphology and geographic range, except that the distribution of subsp. *marginatum* and subsp. *debile* completely overlap (Figs 1, 2). Since Bennett's (1995) description, two additional subspecies have been proposed based on morphological variation such as leaf colour, leaf size, leaf shape and/or growth habit: subsp. '*Busselton*' and subsp. '*Whicher*' (Bennett 2019; A. Webb *pers comm.*) (Figs 1,2); but these are not yet formally described.

The proposed subsp. '*Busselton*' is under threat from impending disturbance within the City of Busselton. It is known from only five populations, four of which occur within a highly restricted <3 km^2 area within the City, while a fifth disjunct pair of sub-populations occurs 30 km away in the Naturaliste area near Sugarloaf Road, discovered in July 2019. All City of Busselton populations grow in areas without formal conservation protection and are subject to disturbance activities via the planned widening of the Vasse Diversion Drain and future expansion of the Busselton Bypass Highway. Both activities would permanently eliminate a significant proportion all known plants. One of the disjunct Naturaliste populations (Three Bears) is the only one to occur within a protected area (Leeuwin-Naturaliste National Park). If subsp. '*Busselton'* is considered a distinct taxon, it would likely fit criteria for Critically Endangered following planned disturbance activities, requiring the implementation of conservation management to offset disturbance to affected natural populations.

Subspecies are often defined as taxa that form geographically and morphologically distinct subgroups within species. The proposed subsp. '*Busselton*' fits this definition within *C. caeruleum*, but formal subspecies nomination is difficult since there is confusion surrounding the wider taxonomy within *C. caeruleum*. For example, recent inspection of Herbarium specimens and field observations have revealed multiple morphological sources of ambiguity such as: subsp. *spathulatum* and subsp. *marginatum* have been mislabelled and confused in Herbarium collections multiple times (Bennett 2019); some Herbarium specimens accessioned as subsp. *oblanceolatum* do not meet diagnostic criteria for subsp. *oblanceolatum* based on Bennett (1995)'s taxonomic key; field observations reveal

strong and inconsistent variation in leaf morphology within and among populations of subsp. *caeruleum* and subsp. *oblanceolatum*; and field observations of subsp. *debile* and subsp. *marginatum*, which overlap in distribution and occur close to the range of subsp. '*Busselton*', indicate highly variable and inconsistent leaf morphology where multiple plants have now been observed to exhibit 'odd' or 'intermediate' leaf morphology, particularly in the southern portion of the range (A. Webb *pers comm*.).

Taxonomic confusion based on strong morphological variation is a larger issue beyond *C. caeruleum*, where wide variation in leaf and growth-form occurs throughout the *Conospermum* genus (Bennett 1995). Geographic and age-related variation in leaf traits has been noted within and among multiple *Conospermum* species, despite the taxonomy relying heavily on leaf characters. Sinclair *et al.* (2008) also noted uncertainty regarding the taxonomy of *C. triplinervium* and found that variable leaf morphology occurred both within and among populations that was suggested to be age-related. In a study of *C. undulatum*, Close *et al.* (2006) also found some evidence that leaf morphological variation was not related to population genetic differentiation, and both of these studies also queried the potential contribution of hybridisation.

Due to such morphological sources of ambiguity, genetic analyses will be required to complement and further improve our understanding of population and phylogenetic relationships of *C. caeruleum*. New and relatively affordable genomic sequencing technologies such as DArTseq™ (which is a 'reduced representation sequencing' method) allows screening of genetic variation across a much larger proportion of the genome (e.g. thousands to tens of thousands of SNP loci) than older genetic technologies previously allowed. Such technologies improve the resolution of both population genetic and phylogenetic analyses of relationships within relatively unknown and closely-related taxa of conservation interest.

1.1 Objectives

The objective of this study is to determine whether *C. caeruleum* subsp. '*Busselton*' shows evidence of genetic differentiation that warrants conservation protection as an independent taxonomic, evolutionary or management unit within the *C. caeruleum* species complex. Using DArTseq, we conduct a SNP-based phylogenomic and population genetic analysis of the *C. caeruleum* species complex to clarify subspecies relationships within the wider group, to provide context for interpreting the status of subsp. '*Busselton*'.

2. METHODS

2.1 Study system

Conospermum caeruleum R.Br. is a species complex belonging to Section *Paniculata*, Subgenus *Conospermum,* within the Australian 'Smoke Bush' genus *Conospermum* (Bennett 1995) (Family Proteaceae) (Weston & Barker 2006). The species occupies three broad geographic areas: (1) the Eastern Jarrah Forest with some extension into the western Wheatbelt (subsp. *spathulatum*); (2) the Albany (subsp. *caeruleum*) and Stirling Range (subsp. *oblanceolatum*) region; and (3) the south west region, on the Swan and Scott coastal plains between Busselton and Scott River, with some extension onto the lateritic plateau between them (subsp. *debile*, subsp. *marginatum*, subsp. '*Busselton*', subsp. '*Whicher*') (Figs 1, 2; Appendix 1). Leaf traits and distributional range define the existing, formally described subspecies (Bennett 1995), but growth form also varies among them: subsp. '*Busselton*', subsp. *debile* and subsp. *marginatum* are prostrate; but subsp. '*Whicher*', subsp. *spathulatum* and subsp. *caeruleum* are erect. Further, subspecies '*Busselton*' and subsp. *caeruleum* can form very large carpets or clumps of plants extending several metres, making it difficult to distinguish discrete individuals; whereas subsp. *debile*, subsp. *marginatum* and subsp. *spathulatum* typically grow as smaller populations of scattered or a few relatively small individual plants (although this can occur in populations of all subspecies). There is ambiguity surrounding the defining morphological characters of subsp. *oblanceolatum* as described by Bennett (1995) that are not consistent in field or Herbarium collections.

Subspecies '*Busselton*' differs from other subspecies in its carpet-forming growth form; fleshy, lime-green coloured leaves that are narrow-oblanceolate to narrow-spathulate and widest toward the tip; dense stem leaf internodes (~1 cm); and is restricted to the Spearwood landform growing in grey sand overlying scattered outcropping limestone, within 1.5 to 2 km of the coast (*pers obs*.; A. Webb *pers comm*.; Bennett 2019). It grows in damp sand above swamps but never in inundated areas (Bennett 2019).

Insects are essential for fertilisation of subsp. *caeruleum* (Stone 2003) where the floral styles are actively-triggered when insects (e.g. bees) visit the flowers, resulting in an 'explosive pollination' mechanism (Stone *et al.* 2006). The mating system of *C. caeruleum* is unknown, but both self-compatibility and strict self-incompatibility have been described in other Western Australian *Conospermum* species (Stone *et al.* 2006, Delnevo *et al.* 2019). Plants of subsp. *caeruleum* near Albany

have been reported to have a seeder fire-response strategy (Bowen & Pate 2017), but post-fire re-sprouting from a lignotuber has also been observed in the Ambergate and Margaret River Headwaters populations of subsp. *marginatum* (Bennett 2019, A. Webb *pers comm*.). Lignotuber re-sprouting may occur in other populations but remains undocumented. *Conospermum* has a chromosome number of *n* = 11 and is very likely diploid (Ramsay 1963; Stace *et al.* 1998).

2.2 Sample collection

Leaves from a total of 229 *C. caeruleum* individuals from 1-5 representative populations of each subspecies were collected in August-September 2019 from the confirmed, currently known range of extant subspecies (Fig 2; Table 1). All fresh subsp. *spathulatum*, subsp. *caeruleum* and subsp. *oblanceolatum* collections were made by M. Byrne and R. Binks. All subsp. *debile*, subsp. *marginatum*, subsp. '*Whicher*' and subsp. '*Busselton*' collections were made by A. Webb, with some assistance from D. Bradbury and R. Binks. Between 6-8 individual leaf samples were collected from each population where possible. At the time of sampling, voucher specimens were collected from all populations and deposited in the Western Australian Herbarium (PERTH). All field collections were identified based on morphology, with the exception of subsp. *oblanceolatum* which was collected based on the distributional range according to Bennett (1995) (populations RGP and SST) due to ambiguity surrounding diagnostic morphological characters. However, DNA was also extracted from one PERTH Herbarium specimen accessioned as subsp. *oblanceolatum* that grows within the expected range of subsp. *caeruleum* (LIT).

A single population on Scott River Road (SCR) in the southern portion of the southwest was highly morphologically variable, being comprised of some individuals with typical subsp. *debile* morphology, some with subsp. *marginatum* morphology, and some with odd and apparently intermediate leaf morphology, referred to as '*intermediate*' (Table 1; Fig 2).

For all samples where possible, leaves were collected from an individual where it could be confirmed that they arose from the same location at ground level, except for subsp. '*Busselton*' and the Betty's Beach (BET) population of subsp. *caeruleum* where individual distinction was more difficult due to the plants growing continuously for up to several metres in some cases. This may be due to vegetative reproduction from an underground lignotuber, or possibly a result of mass-recruitment of multiple individuals from seeds. In these cases, leaf samples were taken over a wide geographic area within

the population, and/or from opposite ends of a large 'patch'. Fresh leaf material was initially dried on silica gel then freeze-dried for long-term storage and DNA extraction.

To provide context for phylogenomic analysis and species delineation, we also sampled two closely related outgroup species from Section *Paniculata* (Bennett 1995) consisting of six individuals from two populations of *C. glumaceum*, and one herbarium specimen of *C. polycephalum* (Table 1).

2.3 DNA extraction and genotyping

DNA was extracted from all samples using a modified 2% CTAB method with the addition of 1% PVP-40 and 0.1% sodium sulphite to the extraction buffer. Due to low DNA concentration, the amount of starting material and extraction buffer was increased (to 60 mg and 800 µL respectively) relative to initial trials to increase yields. Quality of extracted DNA was confirmed by agarose gel electrophoresis. Between 20-50 ng/µL of DNA from a total of 236 individuals was sent to Diversity Arrays Technology (DArT^{TM}) (Canberra) for DArTseq analysis.

Due to the possibility that multiple leaf samples may have been collected from the same genetic individual in some populations, the dataset was screened for replicate genotype samples. A genetic distance matrix was constructed by DArT^{TM}, and all pairwise individual samples with a distance value of less than the average error rate (2%) were considered genetic clones, interpreted as multiple leaf samples taken from the same genetic individual. Replicate genetic samples were removed from the dataset for all downstream analyses.

2.4 DArTseq library preparation, sequencing and filtering

Initial DArTseq library construction was difficult and resulted in many sample failures, which was surprising given the high quality of DNA seen on agarose gel. The material became viscous during DArTseq preparation, which may suggest the presence of unexpected complex polysaccharides co-eluted with DNA. Reactions were improved following sample purification. Library preparation, quality control (including the use of technical replicates to measure reproducibility and Mendelian inheritance), read parameters and original SNP calling were conducted by DArT^{TM} using their proprietary pipelines. This resulted in a dataset of 50,802 SNP loci for our use. We used the 'dartR' analysis package v 1.1.11 (Gruber *et al.* 2018) in R (R Development Core Team, 2013) to further apply quality control filters for various downstream analyses (see below).

2.5 SNP data analysis

2.5.1 Phylogenetic relationships

To investigate phylogenetic relationships within *C. caeruleum*, a subset of the data was used to infer a maximum-likelihood (ML) phylogenetic tree using IQ-TREE v1.6.10 (Nguyen & al., 2015). We applied the best-fit substitution model according to the Bayesian Information Criterion (HKY+F+ASC+G4) and performed 1000 bootstrap replicates with ultrafast bootstrap approximation (Hoang & al., 2018) and the SH-aLRT test for statistical support. The original set of 50,802 SNP loci was filtered to retain only loci with reproducibility >0.98, call rate >0.80, minor allele frequency >0.02 and frequency of heterozygotes >0.02, and to remove population-level noise, the dataset was reduced to two samples per population (66 individuals, 2783 SNPs). *Conospermum polycephalum* was used as the outgroup. The program FIGTREE was used to visualise the tree and to root it by the *C. polycephalum* individual. Only clades with > 95% ultrafast bootstrap support values and > 80% SH-aLRT values were considered to have good support as recommended by the program authors.

2.5.2 Population genetic structure

To investigate population genetic structuring within the *C. caeruleum* species complex we used two genetic clustering methods, each performed at the individual level and therefore not influenced by sampling location. Firstly, we performed multivariate principal coordinates analysis (PCoA) in R using the 'adegenet' package v 2.0.1 (Jombart 2008). This method represents individual genetic relationships as an ordination and does not rely on the assumptions of an evolutionary model. The original dataset of 50,802 SNPs was filtered to retain only the loci with reproducibility >0.98, call rate >0.90, minor allele frequency >0.02 and frequency of heterozygotes >0.02, resulting in a dataset of 1892 high quality SNP loci with 3.5% missing data. Secondly, we conducted a Bayesian analysis of population structuring based on 213 individuals using the program STRUCTURE (Pritchard *et al.* 2000) to detect *K* genetic clusters. For this analysis, the original DArTseq dataset of 50,802 loci was filtered by removing secondary SNP loci to avoid linkage disequilibrium, and by only retaining loci with reproducibility >0.98, call rate >0.80, minor allele frequency >0.02 and frequency of heterozygotes >0.02, resulting in a dataset of 1489 SNPs and 7.7% missing data. Assuming correlated allele frequencies, we applied a 100,000 burnin period and 100,000 MCMC replications to test *K* values ranging from 1 to 8, with 5 iterations of *K*. The most likely value of *K* was determined using STRUCTURE HARVESTER (Earl & vonHoldt 2012), based on the *Delta K* method (Evanno *et al.* 2005), combined with

inspection of the plot of the mean log likelihood of each *K* value (mean LN(K))) and inspection of individual *Q* membership values in the context of biological relevance (Pritchard *et al.* 2000; Janes *et al.* 2017). Consistency across multiple *K* iterations was assessed using CLUMPP (Jakobsson & Rosenberg 2007) and individual *Q* membership proportions were plotted as a barchart, based on the mean Q-matrix over all runs derived from CLUMPP. Piecharts of the mean cluster membership of each population were overlaid on a geographic map to visualise the spatial geographic arrangement of genetic structure.

To further investigate genetic structuring within the southwest region, a subset of 127 individuals was re-analysed by DArT^{TM}'s proprietary pipelines to call SNPs that were specific only to individuals belonging to subsp. *debile*, subsp. *marginatum*, subsp. '*Whicher'* and subsp. '*Busselton'*, to maximise data resolution and quality in this focal region. SNP filtering was performed on this subset as described above for PCoA which was based on 3940 SNP loci and 3.6% missing data. SNP filtering for STRUCTURE analysis was performed for this subset as described above except that call rate was >0.9, resulting in a dataset of 2374 SNP loci and 3.3% missing data.

2.5.3 Population genetic differentiation and diversity

Population differentiation within *C. caeruleum* was estimated using GENEPOP v 4.7.2 (Rousset 2008), which was used to perform global and pairwise population tests of F_{ST} as a measure of genetic differentiation, based on the dataset of 1489 SNPs. Pairwise F_{ST} was also calculated between the three major clades evident on the PCoA (see Results). Genetic diversity indices were also calculated using GENEPOP, including global and population estimates of the fixation index (inbreeding coefficient) (F_{IS}), and expected (HE) and observed (HO) heterozygosity per population. The package 'hierFSTAT' (Goudet 2005) was used in R to calculate allelic richness (A_R) per population.

3. RESULTS

3.1 Data quality and clones

There was a relatively high amount of variation in the degree of missing data among individuals, where rates were higher in outgroup samples and in subsp. *spathulatum* and subsp. *caeruleum* relative to southwest subspecies. At a distance threshold of 2%, 14 samples were detected as genetic repeats (clones) of existing samples in the dataset and were thus removed from all further analyses. These

occurred in subsp. '*Busselton'* (BYP, LWN, SUG), subsp. *marginatum* (AMB), subsp. *caeruleum* (MAR) and subsp. *oblanceolatum* (RGP) (Appendix 2), resulting in 3-7 genetically unique individuals remaining in the analysis for these populations (Table 1). At BYP, clonal samples may have been collected due to the spreading nature of large clumps, where multiple leaf samples were unknowingly collected from the same individual. At the AMB, SUG and LWN populations, plants were scattered but growing in clumps, therefore it is possible that multiple samples were collected from within a single clonal clump. At MAR, plants were growing in and under dense vegetation making identification of discrete individuals at ground level difficult. Due to a very high level of inbreeding in the dataset (see below), it is possible that some very closely related individuals were misidentified as being clonal samples of a single individual, and *vice versa*. However, we consider the rate of these potential errors to be low.

3.2 Phylogenetic relationships

The ML tree identified four major clades that generally reflected geographic regions: (a) all samples of subsp. *spathulatum* (Eastern Jarrah Forest region); (b) all samples of subsp. *caeruleum* and subsp. *oblanceolatum* (Albany and Stirling region); (c) all samples of subsp. *debile* and subsp. *marginatum* collected from the southern southwest (Scott River) region; and (d) all individuals of subsp. *debile*, subsp. *marginatum*, subsp. '*Busselton'* and subsp. '*Whicher'* from the northern southwest region (Fig 3). All four major clades had strong statistical support, while multiple derived clades had poor support. Within the 'SW north' clade, there was good support for a further distinct clade formed by the Ambergate and Margaret Road East (Headwaters) populations. All subsp. '*Busselton*' samples formed a single, strongly supported clade that was nested in a derived position within the 'SW north' clade.

3.3 Population genetic structure

The PCoA of all *C. caeruleum* samples (Fig 4) revealed three major and strongly distinct clusters on the first and second axes, representing: (a) subsp. *spathulatum* (Eastern Jarrah Forest); (b) subsp. *caeruleum* and subsp. *oblanceolatum* (Albany & Stirlings); and (c) all samples of subsp. *debile*, subsp. *marginatum*, subsp. '*Busselton'* and subsp. '*Whicher'* (the southwest). The third axis identified a fourth distinct cluster within the southwest region, equivalent to the 'SW south' from ML analysis. All three axes explained a large proportion (50.9% combined) of the data. Within the Albany and Stirlings cluster, there was mild distinction of populations from the Stirlings region (subsp. *oblanceolatum*) relative to those from the Albany region (subsp. *caeruleum*) (Appendix 3). Within the subsp. *spathulatum* cluster, there did not appear to be any evidence of population structure. All individuals

of subsp. '*Busselton*' clustered together as a discrete group on the PCoA, but with only mild distance from other 'SW north' individuals.

The PCoA of the re-analysed southwest subset (Fig 5) revealed three clearly distinct clusters of individuals on the first two axes corresponding to: (a) SW south; (b) SW north and (c) subsp. '*Busselton*'. The subsp. '*Busselton*' and 'SW south' individuals were most distant from each other, with 'SW north' individuals lying between them, similar to their geographical distribution. There was no apparent population structuring within 'SW south' (Appendix 4). On the third axis, individuals belonging to the Ambergate and Margaret Road East populations were further distinguished. The three axes together explained 38.2% of the variation, with the vast majority of variation captured by PC1.

Bayesian STRUCTURE analysis of all *C. caeruleum* individuals (Figs 6, 7) revealed that *K*=4 was the most likely number of unique genetic clusters, corresponding to: (a) subsp. *spathulatum*; (b) subsp. *caeruleum* and subsp. *oblanceolatum*; (c) 'SW south', (d) subsp. '*Busselton*' + 'SW north'. Within the latter cluster, all subsp. '*Busselton*' individuals showed very strong cluster membership ($Q > 0.99$), but all other 'SW north' individuals showed clear admixture with the 'SW south' cluster (Q range: 0.57 – 0.94). Individuals within all other clusters showed very strong cluster membership with almost no signal of admixture. Individual membership proportions didn't make biological sense at *K*=3. At *K*=5, the 'SW north' individuals formed their own unique cluster but retained a signal of admixture with subsp. '*Busselton*'. At *K*=6, the Ambergate and Margaret Road East populations formed a further unique cluster.

STRUCTURE analysis of the southwest subset (Appendix 5) revealed that *K*=2 was the most likely number of unique genetic clusters based on the Delta K method, but the mean LNP(K) plot did not clearly plateau at *K*=2. This suggests a more rigorous analysis, perhaps with a higher number of MCMC repetitions and burnin period, should be implemented in future analyses of this subset. At K=2, results were almost identical with the analysis of all *C. caeruleum* individuals, where the two clusters corresponded to: (a) 'SW south' with almost complete *Q* membership and no admixture; and (b) subsp. '*Busselton*' and 'SW north' individuals, with almost complete membership of all Busselton individuals, but admixture in all 'SW north' individuals.

3.4 Population genetic differentiation and diversity

The global estimate of population differentiation within the sampled populations of *C. caeruleum* was high (Global F_{ST} = 0.534) representing strong overall genetic structuring but may also be inflated due to high levels of inbreeding (see below). Pairwise population levels of F_{ST} ranged from very low (F_{ST} = 0.018, Geographe Rec Centre *vs* Vasse Diversion Drain) to extremely high (F_{ST} = 0.783, Red Gum Pass *vs* Three Bears), with most values being high (F_{ST}>0.20) (Fig 10).

Pairwise F_{ST} between the 'Albany & Stirlings' cluster and the 'Forest' cluster evident from PCoA was high (0.531). Similarly, pairwise F_{ST} between each of these clusters and the 'southwest' cluster were 0.450 and 0.440, respectively. Within the 'Albany & Stirlings' cluster, mean pairwise population F_{ST} was 0.212 ± 0.02, representing population structuring. Within the 'forest' cluster, mean F_{ST} was much lower at 0.096 ± 0.01. Within the 'southwest' cluster, mean F_{ST} was 0.293 ± 0.01. Within subsp. '*Busselton*', mean pairwise F_{ST} was 0.217 ± 0.04, which was mostly driven by the very high F_{ST} values between the disjunct Naturaliste populations and the Busselton populations. The four City of Busselton populations all shared low pairwise F_{ST} values (< 0.08), indicating close relatedness.

The global estimate of fixation was high (F_{IS} = 0.256) indicating a high level of homozygosity in the dataset, most likely due to significant inbreeding. Population level estimates of observed (H_O) and expected (H_E) heterozygosity are presented in Table 2, while relative levels of population genetic diversity (represented by A_R) and inbreeding (represented by F_{IS}) are presented in Figures 8 and 9. Note that in some populations the total number of individuals is low and therefore diversity estimates may not be as accurate. Mean levels of diversity and fixation within the subsp. *caeruleum*/subsp. *oblanceolatum* (Albany & Stirlings) cluster, and the subsp. *spathulatum* cluster were approximately equal, and were similar to mean diversity in the southwest overall (Fig 11). However, splitting the southwest into its northern (excluding Busselton) and southern parts, the 'SW north' group showed the highest level of diversity, and 'SW south' much lower, comparable with that of subsp. *Busselton*.

The disjunct subsp. '*Busselton*' populations of Sugarloaf Road and Leeuwin National Park (Three Bears) were the least diverse in the dataset, and the Sugarloaf Junction, Busselton Bypass and Vasse Diversion Drain populations were the most homozygous/inbred in the dataset. Genetic diversity within the City of Busselton populations (excluding Naturaliste) was equivalent to or slightly higher than the 'SW south' cluster, but somewhat lower than all other groups.

4. DISCUSSION AND CONCLUSIONS

4.1 Broad Patterns

Based on morphological, geographic and strong genetic differentiation, we suggest that subsp. *spathulatum* should be elevated to species level. There is little population genetic differentiation within this taxon, which was unexpected given the very small population sizes and sparse, scattered plant distribution, together with the ecogeographical disjunction of the Strathmore Hill population. The latter population was restricted to growing on a gravel track along a fenceline bordering the neighbouring nature reserve, but was not seen growing within it. These plants may therefore have been brought to this area via gravel from forest populations when the track was created. Based on field observations, there was limited morphological variation within and among populations of this taxon.

Based on morphological, geographic and SNP genetic data, we also suggest that subsp. *caeruleum* and subsp. *oblanceolatum* should be collapsed into a single, morphologically variable entity, which is elevated to species level. We consider the more substantial genetic differentiation within this group (relative to within subsp. *spathulatum*) to represent population structure. Based on field observations, leaf morphology was highly variable both within and among populations of this group, and sometimes even on a single plant. We were unable to characterise any consistent morphological features that distinguished subsp. *oblanceolatum* from subsp. *caeruleum* in the field, although this was not explicitly tested. The genetic data suggest that population genetic differentiation within this taxon is possibly distributed along an east-west cline, which could be investigated further via STRUCTURE analysis within this group.

Patterns within the third, southwest region were less clear. Based on current genetic and morphological data, we suggest the most parsimonious conclusion is that populations in the southwest represent at least one single, highly morphologically variable species, which is comprised of three management units (MUs). We discuss the details of our reasons for this conclusion below.

4.2. Southwest patterns

Within the southwest, the patterns of genetic differentiation generally mirrored geographic distribution, with some exceptions.

4.2.1. Southwest (South)

The southern ('SW south') group of populations (Scott River Road, Beenup, Govenor-Broome Road, Dennis Road, Milyeannup Hall) consistently formed their own single genetic cluster in multiple genetic analyses (ML, PCoA and STRUCTURE). They are, therefore, relatively closely related individuals that, as a group, are strongly divergent from all other plants in the northern region of the southwest. This suggests that the natural distributional gap between the northern and southern populations is an effective barrier to gene flow, allowing genetic differences to accumulate between regions. The southern group also exhibits relatively low genetic diversity and moderate inbreeding. Despite their genetic similarity with each other and lack of obvious population genetic structure, these plants displayed highly variable morphology within and among populations, including morphology that appeared to be odd and 'intermediate' between subsp. *debile* and subsp. *marginatum*. There are no known morphological characters that can currently define this genetic group.

*4.2.2. Southwest (North), and subsp. '*Busselton*'*

The genetic distinction between subsp. '*Busselton*' populations and all other populations in the northern part of the southwest ('SW north') was less consistent than that of 'SW south'. In the PCoA analysis of the whole dataset, and particularly the PCoA of the southwest subset, individuals of subsp. '*Busselton'* were clearly distinct and formed their own genetic cluster. This was somewhat independent of geography since the Naturaliste populations clustered tightly with City of Busselton populations despite the Ruabon Road, Ambergate NR, and Taylors NR populations occurring in closer geographic proximity (although in a different ecological community). The subsp. '*Busselton*' cluster was distinguishable from the other 'SW north' individuals, which formed a separate group.

The subsp. '*Busselton*' individuals also formed their own strongly supported clade in the ML phylogenetic tree. However, this clade was nested within the larger 'SW north' clade, rather than being an independent lineage. Similarly, in STRUCTURE analysis, the subsp. '*Busselton*' populations were considered to belong to the same genetic cluster as all other 'SW north' individuals. However, the cluster membership of all subsp. '*Busselton*' individuals was almost 100%, while the other 'SW north' individuals were all clearly admixed and therefore showed a consistently different signal of ancestry, despite belonging to the same cluster.

Due to the geographic disjunction between the northern and southern populations within the southwest region, it is unlikely that the signal of admixture in the 'SW north' individuals with 'SW south' is due to contemporary gene flow. It therefore more likely represents shared ancestry. Similarly, the signal of admixture between 'SW north' and subsp. '*Busselton*' also likely reflects shared ancestry, but possibly also reflects contemporary gene flow given their relative geographic proximity. Given the generally low levels of diversity and high levels of fixation/inbreeding in the Busselton populations, however, substantial gene flow with the genetically diverse 'SW north' populations is unlikely.

The data suggest that the subsp. '*Busselton*' populations may be currently evolving on a different evolutionary trajectory from the SW north populations. It is possible that they are adapting to the different ecological environment they currently occupy. Alternatively, their divergence from the 'SW north' may be due to random genetic drift since Busselton plants have some of the lowest genetic diversity and highest levels of inbreeding in the dataset, as opposed to 'SW north' plants which had some of the highest levels of diversity and lowest levels of inbreeding in the dataset. It is possible that the more diverse 'SW north' plants represent the ancestral state, from which the Busselton and 'SW south' groups continue to diverge. The low diversity and higher levels of inbreeding in the Busselton and 'SW south' plants probably also reflects their isolation from 'SW north' plants, particularly the Naturaliste populations. However, these explanations for the patterns observed are speculatory at this stage.

Considering the subsp. '*Busselton*' plants in isolation, they would fit criteria for elevation to subspecies due to morphological, ecogeographical and genetic distinction. However, considered in the context of wider genetic patterns in the southwest region, if the Busselton cluster is elevated to subspecies based on the genetic differentiation reported here, then the same rule should arguably be applied to the 'SW north' group, and certainly to the more genetically differentiated 'SW south' group. However, unlike the Busselton plants, there are no morphological characters that currently define or warrant recognition at subspecies level for the latter two genetic groups. Therefore, our conservative approach is to recognise three conservation genetic Management Units, one of which has morphological support (Busselton), the other two which do not. Further taxonomic and morphological revision may (or may not) identify such characters in the 'SW north' and 'SW south' groups. This may be challenging in a morphologically variable group but we suggest it is the most important next step in future research.

4.3 Incidental findings

Based on the above results, we found no genetic evidence to support the distinction of plants currently considered to be subsp. *debile* versus subsp. *marginatum* as distinguishable taxa. Their genetic relationships largely represented geographic distribution regardless of, and in contrast to, the current taxonomy. There was also no evidence that the population on Sues Road considered to be subsp. '*Whicher*' is a genetically independent taxon, despite this population showing an erect rather than prostrate growth form.

The reason for the strong leaf and growth form morphological variation in this group is currently unknown but may be due to phenotypic plasticity influenced by unknown environmental factors, or a function of plant age (or both). Regardless, current assessments of leaf and growth form morphology on which the taxonomy relies often do not correspond to patterns of SNP genetic differentiation in the 'southwest' taxon/taxa of *C. caeruleum* observed here.

We found a consistent relationship between the Ambergate and Margaret Road East (Headwaters) populations within the 'SW north' group. This was unexpected because these populations are not geographically proximate and to our knowledge do not share any obvious or uniquely defining morphological features. However, the Ambergate population was noted to be resprouting from a lignotuber following recent fire (A. Webb pers. comm.), and a population of the Margaret River Headwaters was also described by E. Bennett to have an obvious lignotuber (Bennett 2019). However, it is unknown if this trait is unique to these populations or shared by others in the group as it has not been explicitly recorded.

4.4. Summary and Recommendation

Based on phylogenetic and population genetic analysis, we suggest that there are at least three related but distinct *Conospermum* species within the dataset that mirror the three broad geographic regions assessed in this study: (1) subsp. *spathulatum* cluster (Forest region); (2) subsp. *caeruleum* and subsp. *oblanceolatum* cluster (Albany & Stirlings region), and (3) the 'southwest' cluster, including subsp. *debile*, subsp. *marginatum*, subsp. '*Whicher'* and subsp. '*Busselton'*. Our conservative approach is to consider subsp. '*Busselton'* as one of three independent Management Units (MUs) that belong to a single, morphologically variable southwest species, which exhibits strong population genetic structure. We consider this preliminary conclusion to be the most parsimonious and conservative

relative to alternatives. We suggest that further research is required to better understand the evolution of genetic diversity and differentiation in this group, and to support or reject our conclusion.

5. FURTHER RESEARCH

The following research could be considered to provide further information on this species:

- A thorough morphological and taxonomic revision of the group including all specimens collected as part of this study.
- Further molecular study of the chloroplast genome may provide additional insight to historical evolutionary and demographic processes, patterns of seed dispersal and levels of divergence across the whole *Conospermum caeruleum* complex, but particularly within the SW region.
- A study of adaptive genomic divergence via F_{ST} outlier and/or environmental association analyses, to search for any evidence of local adaptation involving the '*Busselton'* MU.
- Consideration of the influence of, or evidence of, contemporary or historical hybridisation with other co-occurring *Conospermum* species in the SW region. The potential for hybridisation in other *Conospermum* ranges from the ability to produce fertile F1 plants to complete cross-incompatibility, depending on species (Morrison *et al.* 1994; Sinclair *et al.* 2008).
- An investigation of the breeding system to better understand patterns of gene flow, potential pollinator behaviour and seed dispersal, and causes of high levels of inbreeding (e.g. self-fertilisation vs mating among close relatives).

6. ACKNOWLEDGEMENTS

We thank Andrew Webb for performing all collections of subsp. *debile*, subsp. *marginatum*, subsp. *Busselton* and subsp. *Whicher*, and for providing voucher specimens and detailed information regarding morphological variation and population location of multiple subspecies. We thank Mike Hislop for helpful discussion, particularly regarding the morphology of subsp. *caeruleum* and subsp. *oblanceolatum*. We thank Bronwyn McDonald for laboratory assistance.

7. REFERENCES

Bennett EM (1995) Conospermum. In: *Flora of Australia, Vol 16*, pp. 224–271. CSIRO Australia.

Bennett E (2019) *Distribution of Conospermum caeruleum subsp. Busselton*: final report prepared for the Water Corporation, WA. 42p.

Bowen BJ, Pate JS (2017) Patterns of storage tissue and starch distribution in the young taproot of obligate seeders and resprouters of Australian Proteaceae (Juss.): Possible evidence of homoplastic evolution. *Austral Ecology*, **42**, 617–629.

Delnevo N, van Etten EJ, Byrne M, Stock WD (2019) Floral display and habitat fragmentation: Effects on the reproductive success of the threatened mass - flowering *Conospermum undulatum* (Proteaceae). *Ecology and Evolution*, 1–10.

Earl DA, vonHoldt BM (2012) STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genetics Resources*, **4**, 359–361.

Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Molecular Ecology*, **14**, 2611–2620.

Goudet J (2005) HIERFSTAT, a package for R to compute and test hierarchical *F*-statistics. *Molecular Ecology Notes*, **5**, 184–186.

Goudet J, Jombart T (2015) Package 'hierfstat.'

Gruber B, Unmack PJ, Berry OF, Georges A (2018) dartR: An R package to facilitate analysis of SNP data generated from reduced representation genome sequencing. *Molecular Ecology Resources*, 18(3): 691-699 .

Jakobsson M, Rosenberg NA (2007) CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinformatics*, **23**, 1801–1806.

Janes JK, Miller JM, Dupuis JR *et al.* (2017) The K = 2 conundrum. *Molecular Ecology*, **26**, 3594–3602.

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

Morrison DA, McDonald M, Bankoff P, Quirico P, Mackay D (1994) Reproductive isolation mechanisms among four closely-related species of *Conospermum* (Proteaceae). *Biological Journal Of The Linnean Society*, **116**, 13–31.

Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics*, **155**, 945–959.

Ramsay HP (1963) Chromosome numbers in the proteaceae. *Australian Journal of Botany*, **11**, 1–20.

Rousset F (2008) genepop'007: a complete re-implementation of the genepop software for Windows and Linux. *Molecular Ecology Resources*, **8**, 103–106.

Sinclair E, Cheetham B, Krauss S, Hobbs R (2008) Morphological and molecular variation in *Conospermum triplinervium* (Proteaceae), the tree smokebush: Implications for bushland restoration. *Australian Journal of Botany*, **56**, 451–460.

Stace HM, Douglas AW, Sampson JF (1998) Did "paleo-polyploidy" really occur in Proteaceae? *Australian Systematic Botany*, **11**, 613–629.

Stone LM (2003) Floral biology and propagation of blue-flowered *Conospermum* spp. PhD Thesis, Murdoch University.

Stone LM, Seaton KA, Byrne M, Mccomb JA (2006) A study of the reproductive biology of blue-flowered *Conospermum* species (Proteaceae). *Australian Journal of Botany*, **54**, 543–551.

Weston PH, Barker NP (2006) A new suprageneric classification of the Proteaceae, with an annotated checklist of genera. *Telopea*, **11**, 314–344.

Table 1. Collection details, location information and number of unique individuals analysed (*n*) for all sampled populations of *Conospermum caeruleum* and outgroups, for phylogenomic and population genetic analysis.

Species	Subspecies	Geographical region	Population Name	Code	Collection Type	Latitude	Longitude	*n*
C. caeruleum	'*Busselton*'	Busselton	Busselton Bypass Highway	BYP	Fresh	-33.67144	115.33733	4
		Busselton	Vasse Diversion Drain	VDD	Fresh	-33.66489	115.33033	8
		Busselton	Geographe Leisure Centre	GLC	Fresh	-33.66625	115.32653	6
		Busselton	Golf Course (Par 3)	GP3	Fresh	-33.66852	115.31586	8
		Cape Naturaliste	Sugarloaf Junction	SUG	Fresh	-33.55784	115.03752	2
		Cape Naturaliste	Leeuwin-Naturaliste National Park (3 Bears)	LWN	Fresh	-33.55712	115.03195	3
	caeruleum	Albany/Denmark	Nutcracker Road	NUT	Fresh	-34.86225	117.36527	8
		Albany	Narrikup	NRK	Fresh	-34.77146	117.69708	8
		Albany	Marbelup Road	MAR	Fresh	-34.98491	117.72189	7
		Albany	Betty's Beach	BET	Fresh	-34.93791	118.20172	7
		Albany	Little Beach (PERTH 6903223)*	LIT	Herbarium	-34.97200	118.19400	1
	oblanceolatum	Stirlings	Red Gum Pass Road	RGP	Fresh	-34.44219	117.73383	7
		Stirlings	South Stirling	SST	Fresh	-34.56361	118.22376	8
	spathulatum	Eastern Jarrah Forest	Gibbs Road	GIB	Fresh	-33.37634	116.65902	8
		Eastern Jarrah Forest	North Kukilup Nature Reserve	KUK	Fresh	-33.77413	116.68475	8
		Eastern Jarrah Forest	Haddleton Nature Reserve	HAD	Fresh	-33.60462	116.62943	8
		Eastern Jarrah Forest	Bowelling	BOW	Fresh	-33.43566	116.48285	8
		Wheatbelt	Strathmore Hill	STH	Fresh	-33.60758	117.35377	8
	debile	Southwest - north	Ruabon Nature Reserve	RNR	Fresh	-33.64539	115.50961	8
		Southwest – north	Treeton	TRE	Fresh	-33.79860	115.29062	8
		Southwest – north	Taylors	TAY	Fresh	-33.75505	115.20129	8
		Southwest – south	Governor-Broome Road	GBR	Fresh	-34.24896	115.31255	8
		Southwest – south	Dennis Road	DEN	Fresh	-34.24000	115.32344	3
		Southwest – south	Scott River Road**	SCR**	Fresh	-34.19445	115.26749	6
	marginatum	Southwest – north	Ruabon Road (Russ)	RRD	Fresh	-33.64192	115.49017	8
		Southwest – north	Ambergate	AMB	Fresh	-33.73858	115.32275	3
		Southwest – north	Margaret Road East (Headwaters)	MRE	Fresh	-33.89077	115.44024	8
		Southwest – south	Milyeanup Hall	MLY	Fresh	-34.25015	115.44703	8
		Southwest – south	Beenup	BEE	Fresh	-34.22069	115.26891	8
		Southwest – south	Scott River Road**	SCR**	Fresh	-34.19445	115.26749	6
	intermediate	Southwest – south	Scott River Road**	SCR**	Fresh	-34.19445	115.26749	6
	'*Whicher*'	Southwest - north	Sues Road	SUE	Fresh	-33.79356	115.41921	8
C. glumaceum			Boonanarring Nature Reserve	BOO	Fresh	-31.23503	115.87044	3
			Udumung Nature Reserve	UDU	Fresh	-31.18258	116.17594	3
C. polycephalum			Koodjee NR (PERTH 9069208)	KOO	Herbarium	-30.88800	116.02700	1

*This collection is accessioned as subsp. *oblanceolatum* in the PERTH Herbarium, but occurs within the expected range of subsp. *caeruleum*.

**Individuals appearing to have the morphology of subsp. *debile (*SCRd), subsp. *marginatum* (SCRm) and '*intermediate'*(SCRi) leaf morphology all co-occur in this population.

Table 2. Estimates of observed (Ho) and expected (He) heterozygosity within each sampled population of *Conospermum caeruleum* based on 1489 SNPs.

Region	Subspecies	Population	H_O	H_E
Albany	*caeruleum*	Bettys Beach	0.0645	0.0972
Albany	*caeruleum*	Marbelup Rd	0.0848	0.1078
Albany	*caeruleum*	Narrikup	0.0853	0.1065
Albany	*caeruleum*	Nutcracker	0.0823	0.1018
Busselton	*'Busselton'*	Drain	0.0479	0.0937
Busselton	*'Busselton'*	Highway	0.0496	0.088
Busselton	*'Busselton'*	Golf Course	0.0644	0.0946
Busselton	*'Busselton'*	Rec Centre	0.0776	0.0994
Busselton	*'Busselton'*	3Bears	0.0261	0.0396
Busselton	*'Busselton'*	Sugarloaf	0.0179	0.0294
Forest	*spathulatum*	Bowelling	0.0843	0.1088
Forest	*spathulatum*	Gibb Rd	0.0861	0.1076
Forest	*spathulatum*	Haddleton	0.0781	0.1105
Forest	*spathulatum*	North Kukilup	0.0707	0.0954
Forest	*spathulatum*	Strathmore Hill	0.0864	0.0992
Stirlings	*oblanceolatum*	Red Gum Pass	0.058	0.0693
Stirlings	*oblanceolatum*	South Stirling	0.0628	0.0847
SW north	*debile*	Ruabon NR	0.1122	0.1244
SW north	*debile*	Taylors	0.0767	0.1165
SW north	*debile*	Treeton	0.0911	0.1237
SW north	*marginatum*	Ambergate	0.1246	0.1306
SW north	*marginatum*	Margaret Rd East	0.0829	0.098
SW north	*marginatum*	Ruabon Russells	0.0864	0.1124
SW north	*'Whicher'*	Sues Rd	0.081	0.0983
SW south	*debile*	Dennis Rd	0.0606	0.0819
SW south	*debile*	GovBroome Rd	0.0572	0.0849
SW south	*debile*	Scott River deb	0.0549	0.0845
SW south	*'intermediate'*	Scott River inter	0.0604	0.0833
SW south	*marginatum*	Scott River margi	0.0538	0.0871
SW south	*marginatum*	Beenup	0.0621	0.0831
SW south	*marginatum*	Mily Hall	0.0524	0.0773

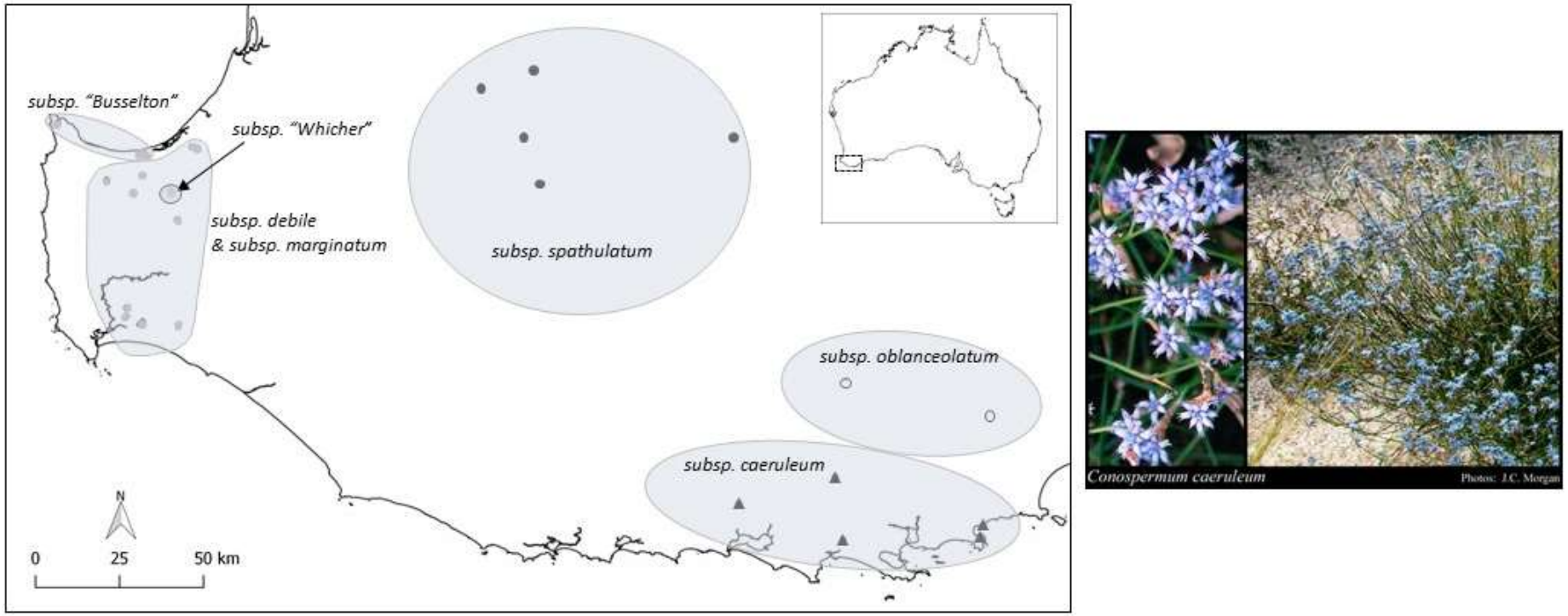


Figure 1. Map of the approximate geographic distribution of currently described and proposed subspecies of *Conospermum caeruleum* (pictured, Florabase), overlaid on populations sampled for the present study (see Table 1, Fig 2 for details).

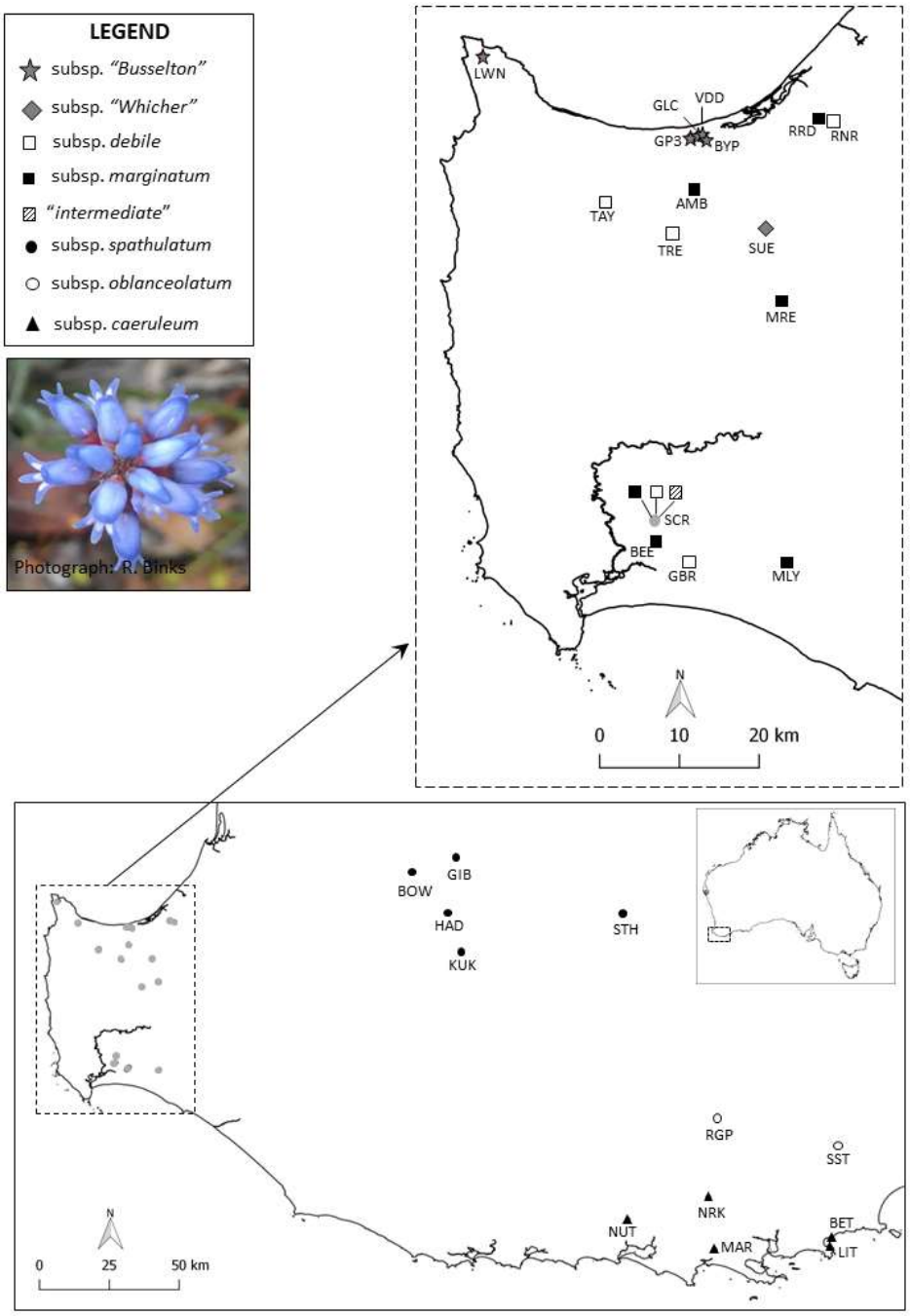


Fig 2. Geographic map of all collection locations of 229 individuals of *Conospermum caeruleum* for SNP phylogenomic analysis, including population name codes (refer to Table 1) and zoomed-in inset of southwest focal area, which contains subsp. '*Busselton*', subsp. '*Whicher*', subsp. *debile* and subsp. *marginatum*. Legend denotes subspecies allocation at the time of collection based on morphology and/or distributional range according to Bennett (1995). Note: Three morphological types co-occur at the single Scott River Road (SCR) population.

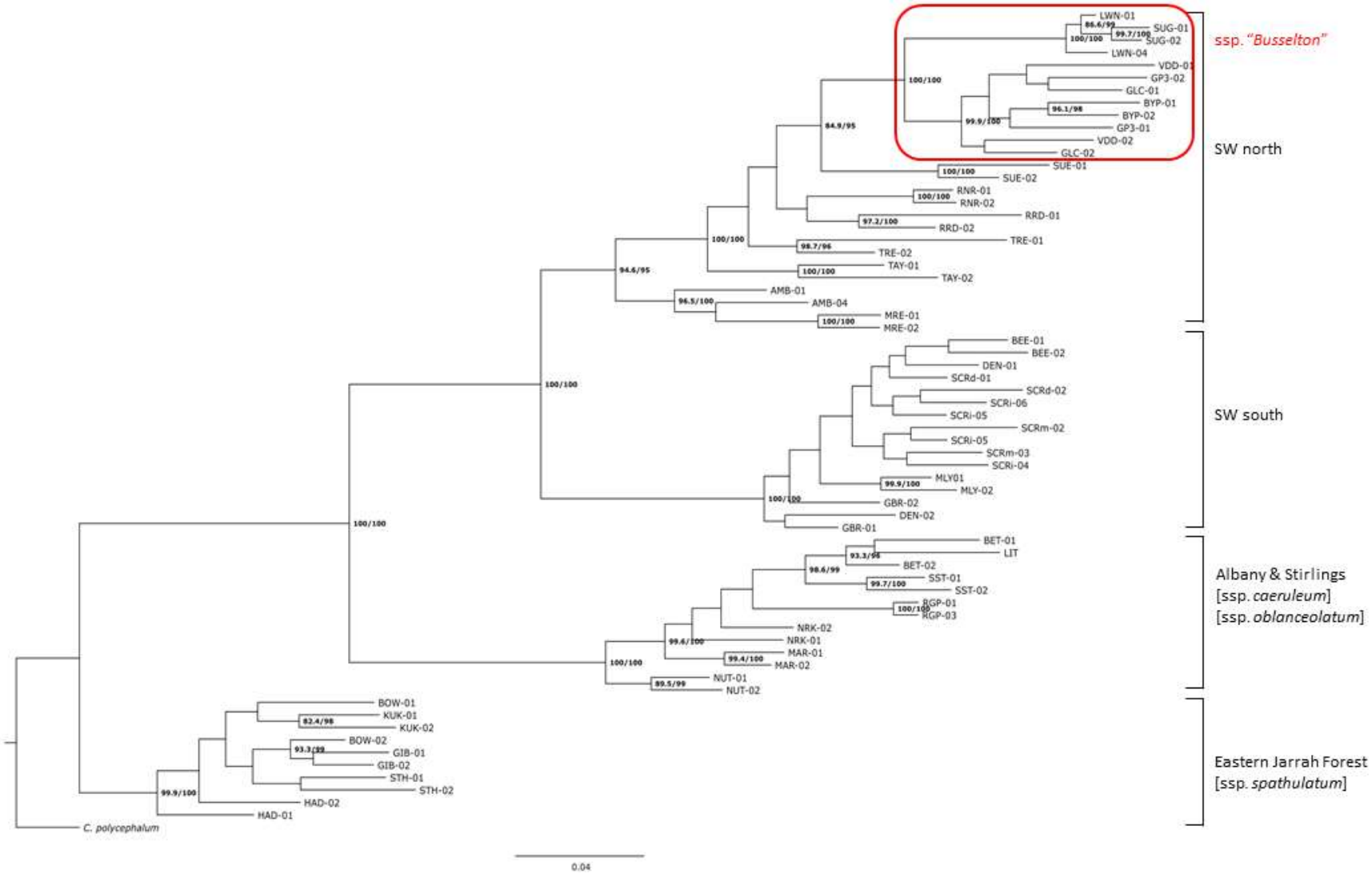


Fig 3. Maximum likelihood (ML) phylogenetic tree of 66 individuals of *Conospermum caeruleum* based on 2783 SNPs, with *C. polycephalum* as outgroup. Refer to Table 1 for population codes. Branch lengths indicate substitutions/site. Values at nodes indicate SH-aLRT (%)/ultrafast bootstrap support (%). Only the support values ≥ 80%/95% are shown.

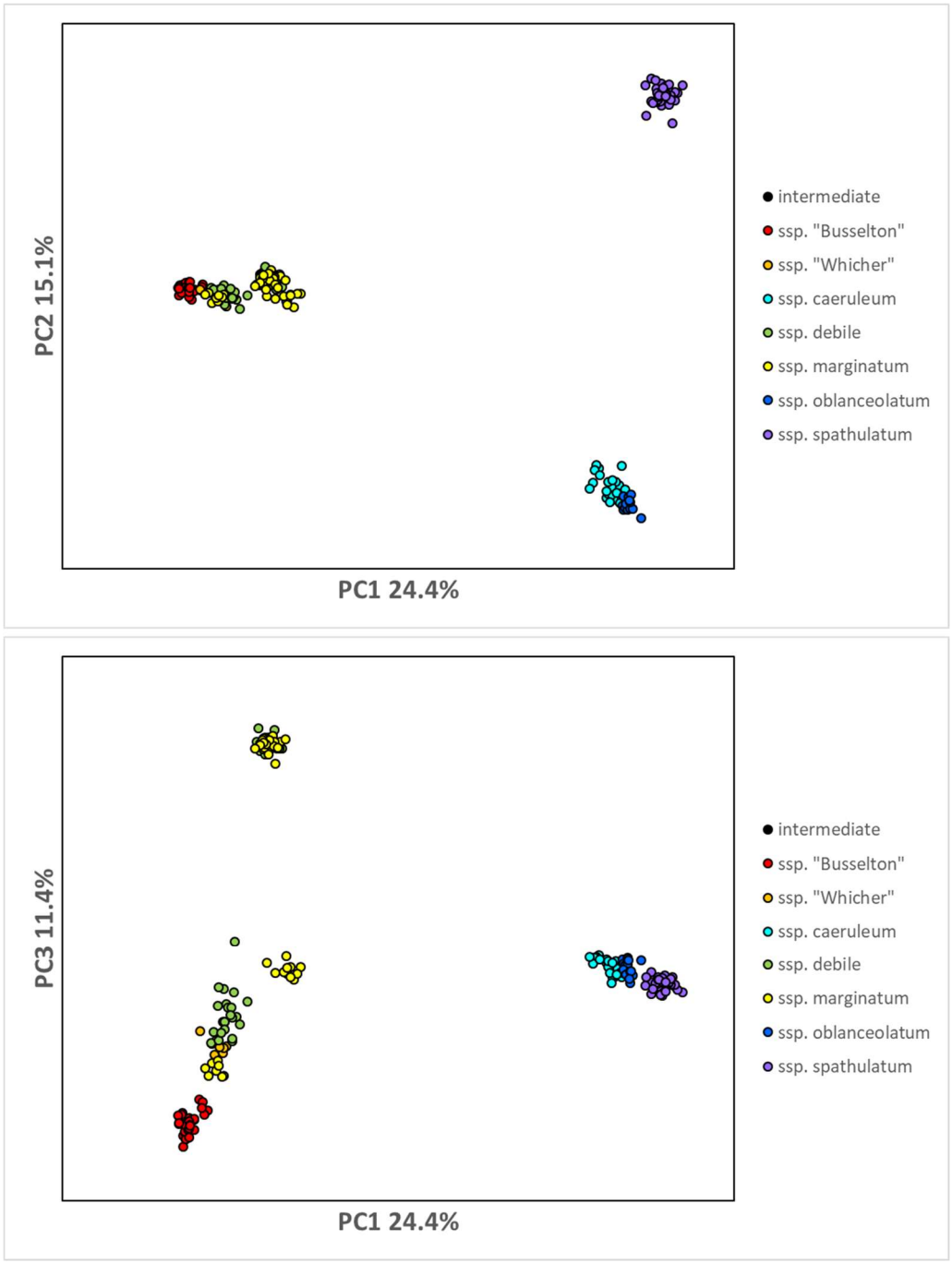


Figure 4: PCoA ordination of 213 individuals of *Conospermum caeruleum* based on 1892 SNPs. Individuals are colour coded by existing subspecies nomination. Upper plot: PC 1 *vs* PC 2; Lower plot: PC 1 *vs* PC 3.

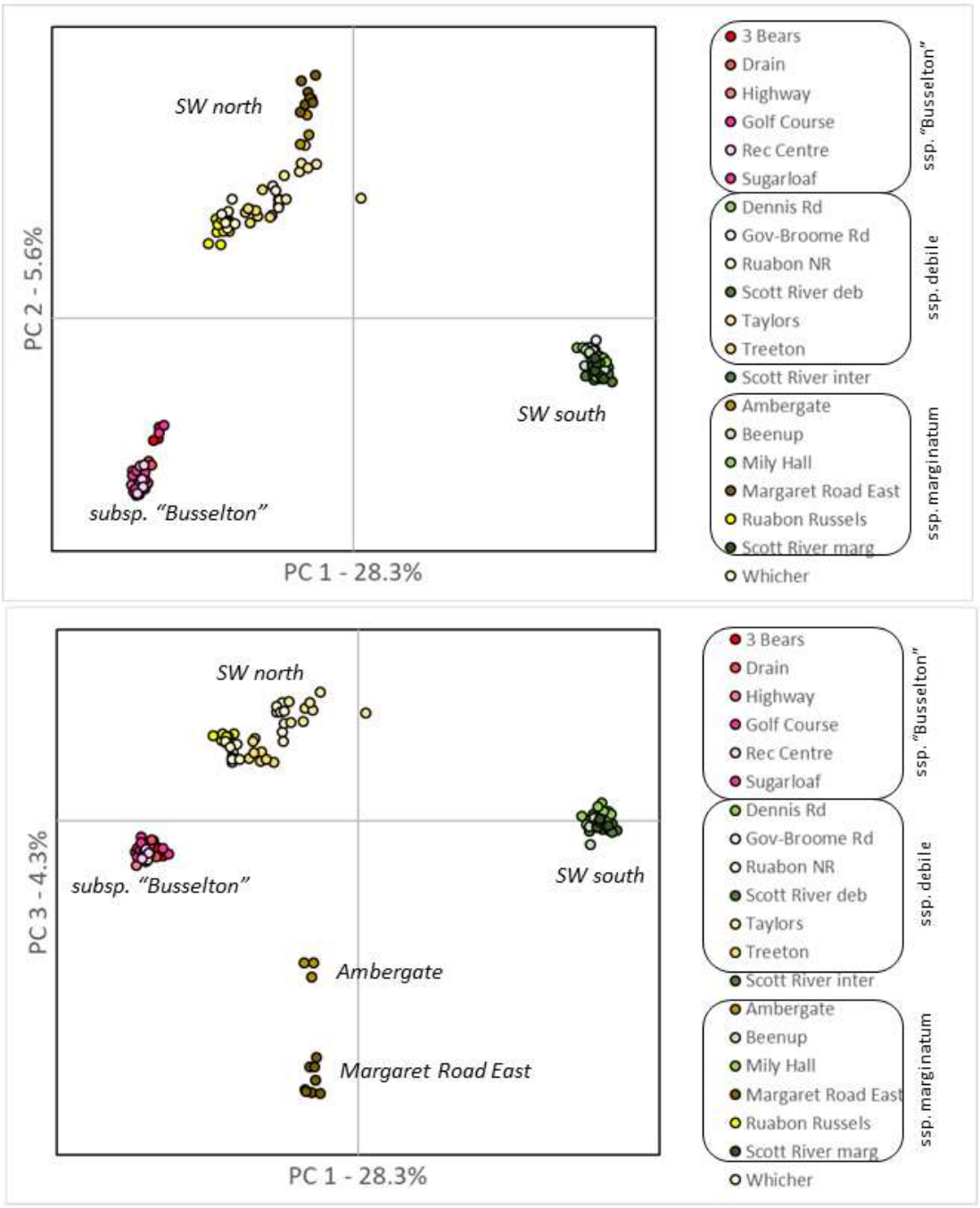


Figure 5. PCoA ordination of 127 individuals of *Conospermum caeruleum* from the southwest region subset, based on 3940 SNPs. Upper plot: PC 1 *vs* PC 2; Lower plot: PC 1 *vs* PC 3.

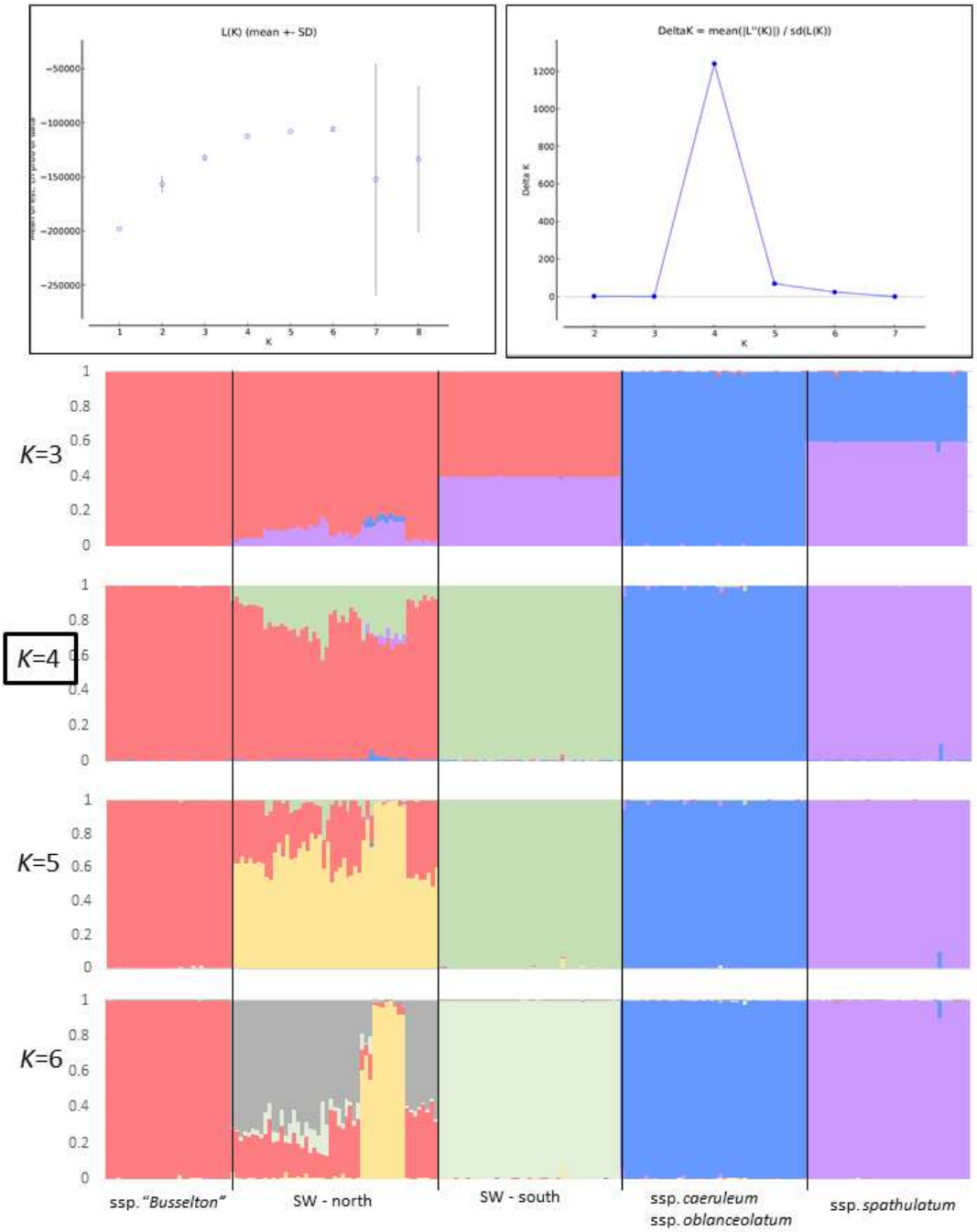


Figure 6. STRUCTURE analysis of 213 individuals of *Conospermum caeruleum* based on 1489 SNPs. Plots of lnP(K) and *Delta K* represent the most likely *K*. Barcharts of individual membership proportions to each cluster are presented below for *K*=2 to K=6. The most likely is *K*=4.

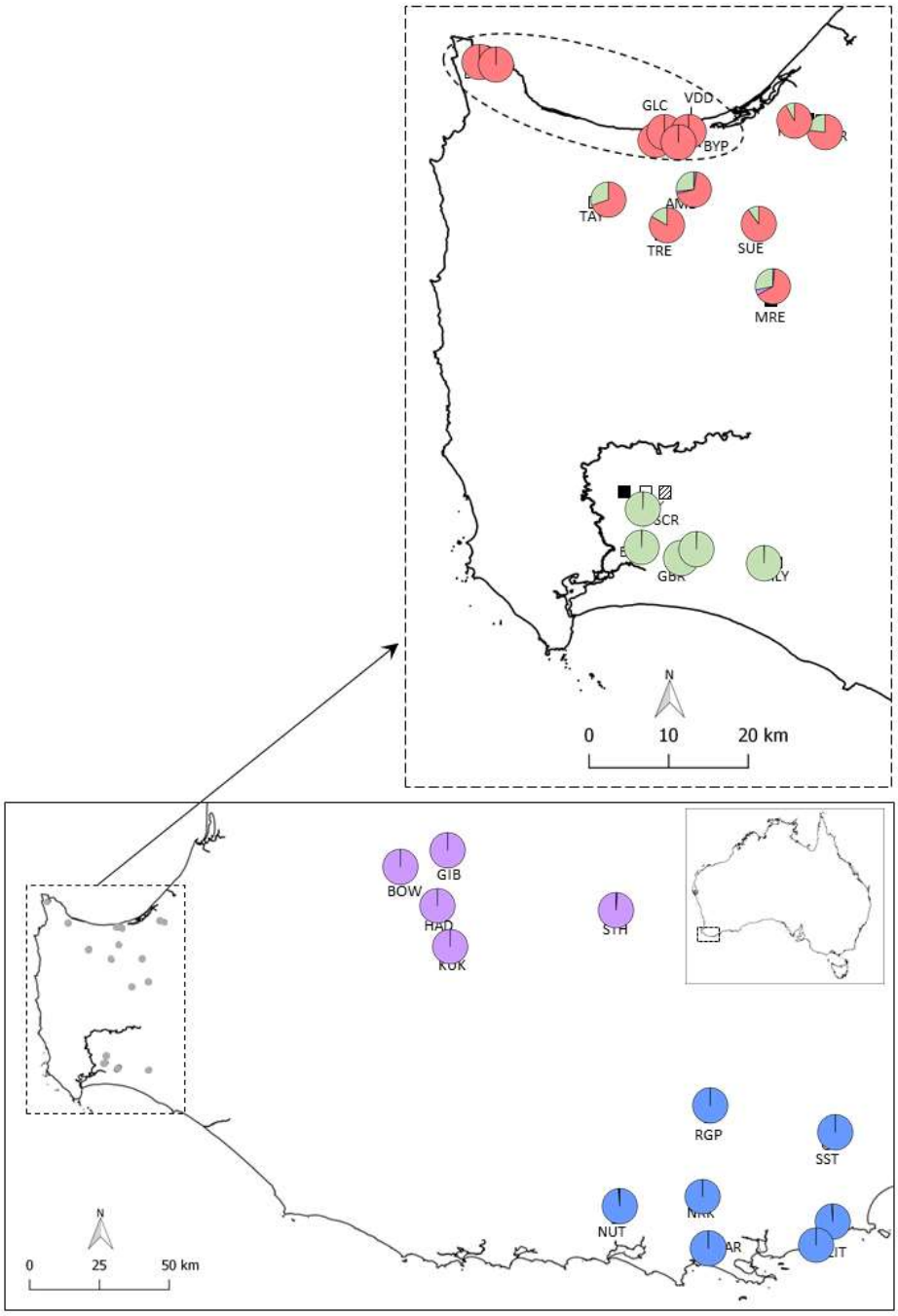


Fig 7. Pie charts of mean cluster membership proportion of each population of *Conospermum caeruleum* when *K*=4, overlaid on a geographic map of sampled populations. Data based on STRUCTURE analysis of 213 individuals using 1489 SNPs. Populations of subsp. '*Busselton*' are circled in the upper inset map.

Genetic Diversity (A_R)

Region	Subspecies	Population	A_R
Busselton	Busselton	Sugarloaf junction	1.024935
Busselton	Busselton	Three Bears	1.035505
Stirlings	oblanceolatum	Red Gum Pass	1.065852
SW south	marginatum	Milyeanup Hall	1.075242
SW south	debile	Dennis Road	1.07622
SW south	marginatum	Beenup	1.080336
SW south	intermediate	Scott River Road inter	1.080342
SW south	debile	Scott River Road deb	1.081061
Stirlings	oblanceolatum	South Stirling NR	1.08148
Busselton	Busselton	Bypass Highway	1.081633
SW south	debile	GovenorBroome Road	1.082129
SW south	marginatum	Scott River Road margi	1.083072
Forest	spathulatum	North Kukilup	1.089505
Busselton	Busselton	Drain	1.089899
Busselton	Busselton	Par 3 Golf Course	1.091251
Albany	caeruleum	Bettys Beach	1.093321
Forest	spathulatum	Strathmore Hill	1.094921
SW north	marginatum	Margaret Road East	1.095893
SW north	Whicher	Sues Road	1.096294
Busselton	Busselton	Geographe Rec Centre	1.096512
Albany	caeruleum	Nutcracker Road	1.0991
Forest	spathulatum	Haddleton	1.101953
Forest	spathulatum	Bowelling	1.102505
Albany	caeruleum	Narrikup	1.103587
Forest	spathulatum	Gibbs Road	1.103649
Albany	caeruleum	Marbelup Road	1.104768
SW north	marginatum	Ruabon Road Russells	1.110164
SW north	debile	Taylors NR	1.113121
SW north	debile	Treeton	1.120845
SW north	debile	Ruabon Nature Reserve	1.123468
SW north	marginatum	Ambergate	1.128735

Least Diverse

Most Diverse

Figure 8. Relative levels of population genetic diversity (based on allelic richness (A_R)), represented by red to green colour scale, in *Conospermum caeruleum* based on 1489 SNPs.

Fixation Index (F_{IS}) (Inbreeding Coefficient)

Region	Subspecies	Population	F_{IS}
SW north	marginatum	Ambergate	0.0462
SW north	debile	Ruabon NR	0.0983
Forest	spathulatum	Strathmore Hill	0.1288
SW north	marginatum	Margaret Rd East	0.1541
Stirlings	oblanceolatum	Red Gum Pass	0.1634
SW north	Whicher	Sues Rd	0.1763
Albany	caeruleum	Nutcracker	0.192
Albany	caeruleum	Narrikup	0.1991
Forest	spathulatum	Gibb Rd	0.1994
Albany	caeruleum	Marbelup Rd	0.2132
Busselton	Busselton	Rec Centre	0.2194
Forest	spathulatum	Bowelling	0.2246
SW north	marginatum	Ruabon Russells	0.2318
SW south	marginatum	Beenup	0.2531
Stirlings	oblanceolatum	South Stirling	0.2588
Forest	spathulatum	North Kukilup	0.2595
SW south	debile	Dennis Rd	0.2598
SW north	debile	Treeton	0.2635
SW south	intermediate	Scott River inter	0.2749
Forest	spathulatum	Haddleton	0.294
Busselton	Busselton	Golf Course	0.319
SW south	marginatum	Mily Hall	0.3225
SW south	debile	GovBroome Rd	0.3263
Albany	caeruleum	Bettys Beach	0.3366
Busselton	Busselton	3Bears	0.3395
SW north	debile	Taylors	0.3414
SW south	debile	Scott River deb	0.3503
SW south	marginatum	Scorr River marg	0.3822
Busselton	Busselton	Sugarloaf	0.3924
Busselton	Busselton	Highway	0.4365
Busselton	Busselton	Drain	0.4883

Least inbred

Most inbred

Figure 9. Relative degree of inbreeding per population (based on the fixation index (F_{IS})), represented by green to red colour scale, in *Conospermum caeruleum*, based on 1489 SNPs.

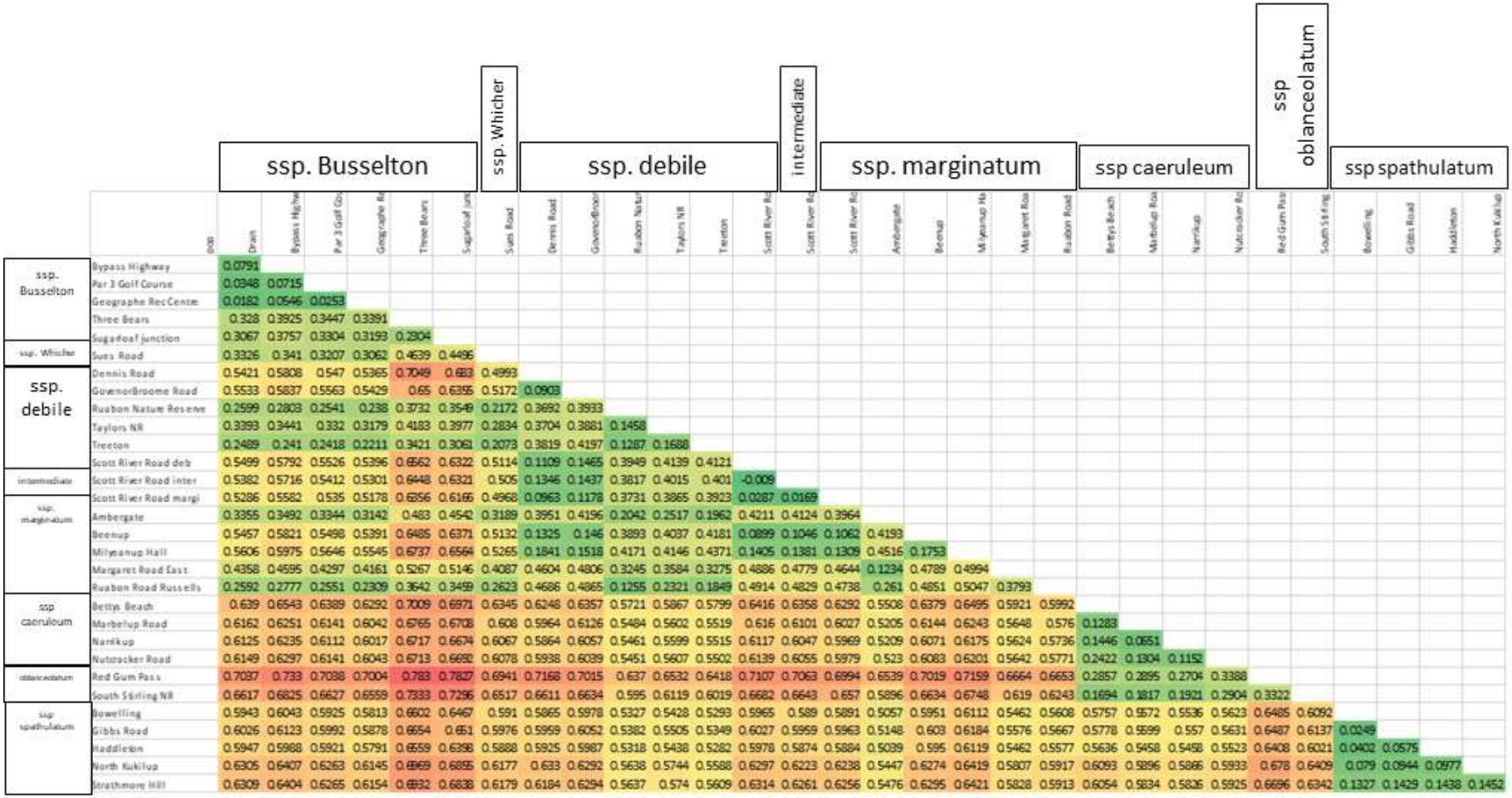


Figure 10. Pairwise F_{ST} values between all sampled populations of *Conospermum caeruleum* based on 1489 SNPs. Colour scale represents relative levels of differentiation (dark green = lowest; dark red = highest).

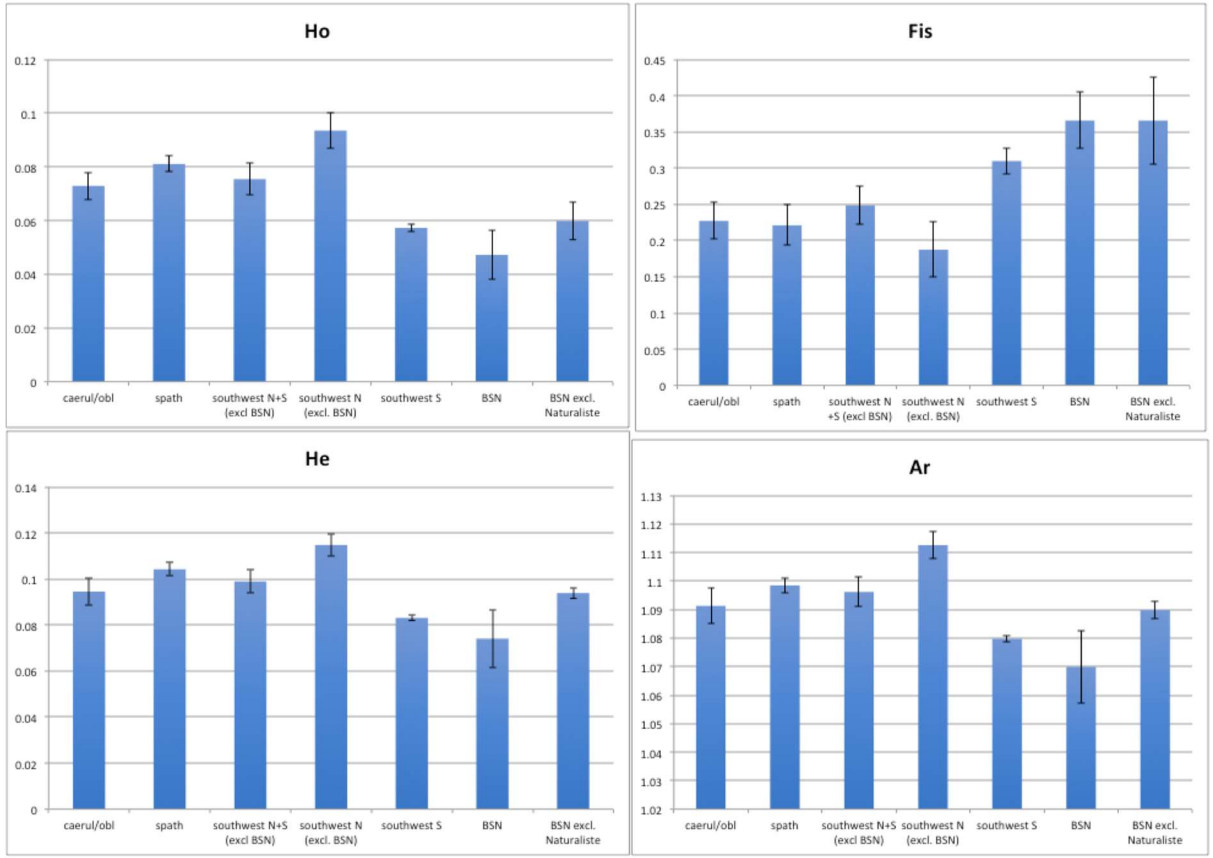
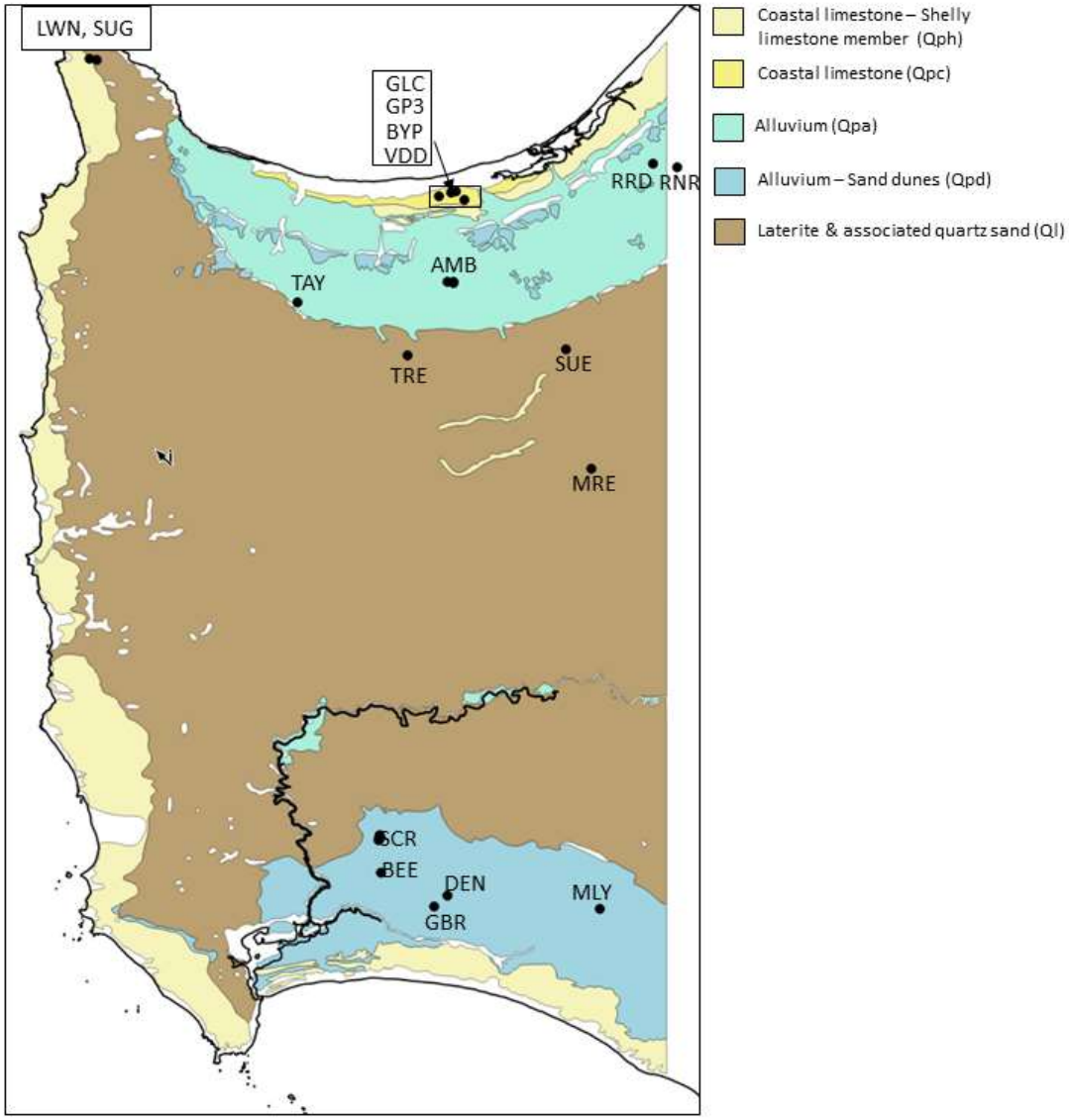


Fig 11. Mean levels of diversity represented by observed heterozygosity (Ho), expected heterozygosity (He), fixation index/inbreeding coefficient (Fis) and allelic richness (Ar) per region and/or genetic cluster of *Conospermum caeruleum*, based on 1489 SNPs.

APPENDIX 1

Geology of the 1:250 000 mapsheet, BUSSELTON-AUGUSTA

(Geological Survey of Western Australia)
Note that only the units relevant to C. caeruleum populations are mapped



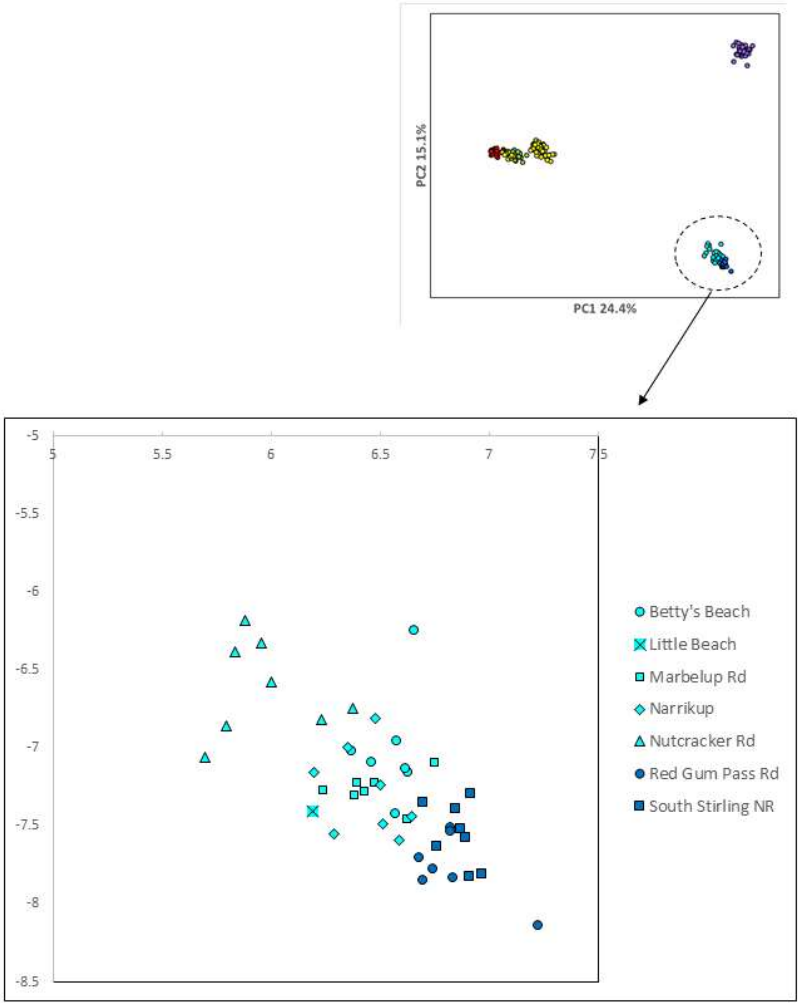
APPENDIX 2

Genetic repeat/clonal samples identified

Subspecies	**Population**	**Samples considered genetic clones**
subsp. '*Busselton*'	Busselton Bypass	A: 02, 03 B: 05, 06, 07
	Leeuwin National Park (Three Bears)	A: 01, 02, 03 B: 05, 06
	Sugarloaf Junction	02, 03
subsp. *marginatum* (SW north)	Ambergate	A: 01, 02, 03 B: 04, 05 C: 06, 07, 08
subsp. *caeruleum*	Marbelup Road	05, 06
subsp. *oblanceolatum*	Red Gum Pass Road	01, 02

APPENDIX 3

Expanded inset of population structure within the Albany & Stirlings (subsp. *caeruleum* + subsp. *oblanceolatum*) cluster from PCoA analysis based on 1892 SNPs.

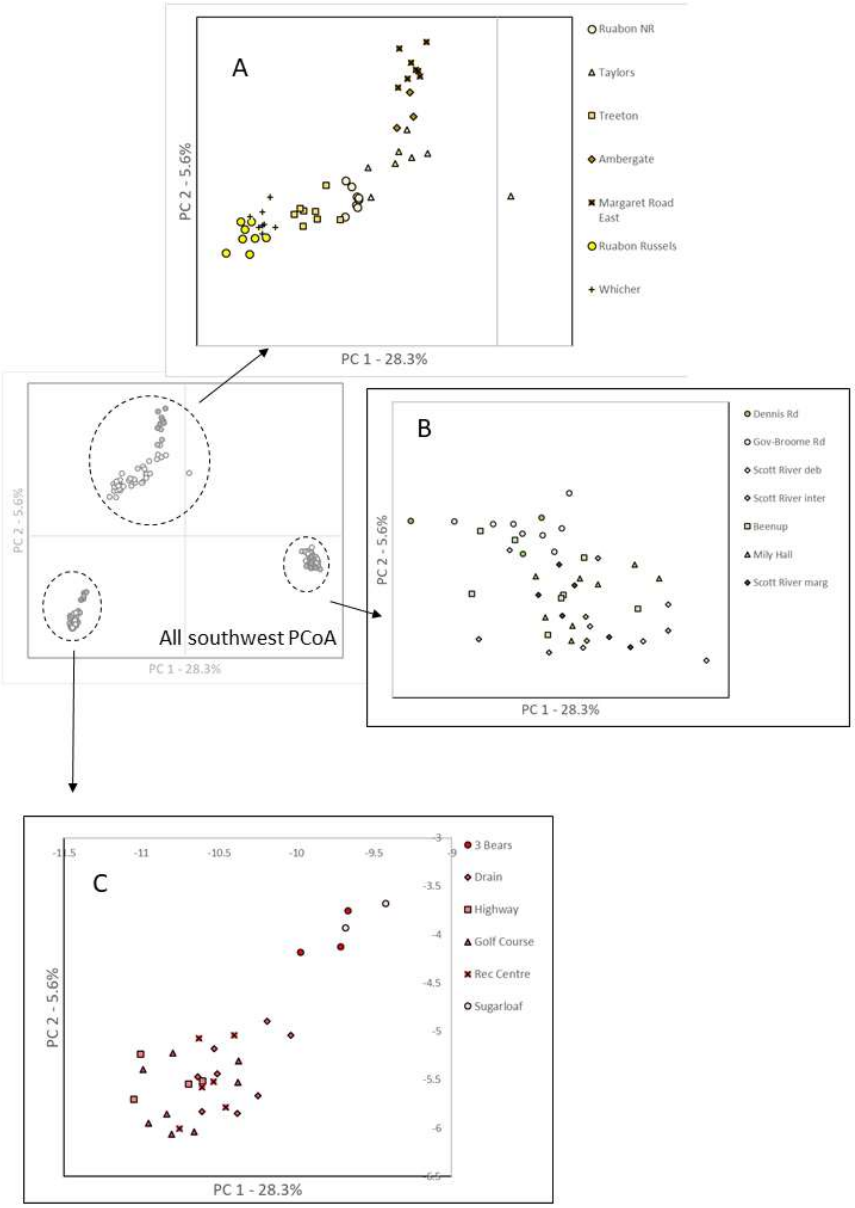


APPENDIX 4

Expanded PCoAs of each separate cluster of the southwest subset PCoA, showing population structure (or lack thereof) within each cluster.

Only PC1 and PC 2 are shown.

A – 'SW north'; B – 'SW south'; C – subsp. '*Busselton*'.



APPENDIX 5

Results of STRUCTURE analysis of the *Conospermum caeruleum* southwest subset based on 2374 SNPs.