SUPPLEMENTAL MATERIAL

Appendix S1. Location and climate details for sampled populations

Appendix S2: Locating the South African source population for east Australian *Arctotheca populifolia*.

Appendix S3: Additional experimental details

Appendix S4: Differences among introduced populations in Australia

Appendix S5: Multivariate analysis of source and introduced plants

Appendix S6: Raw data.

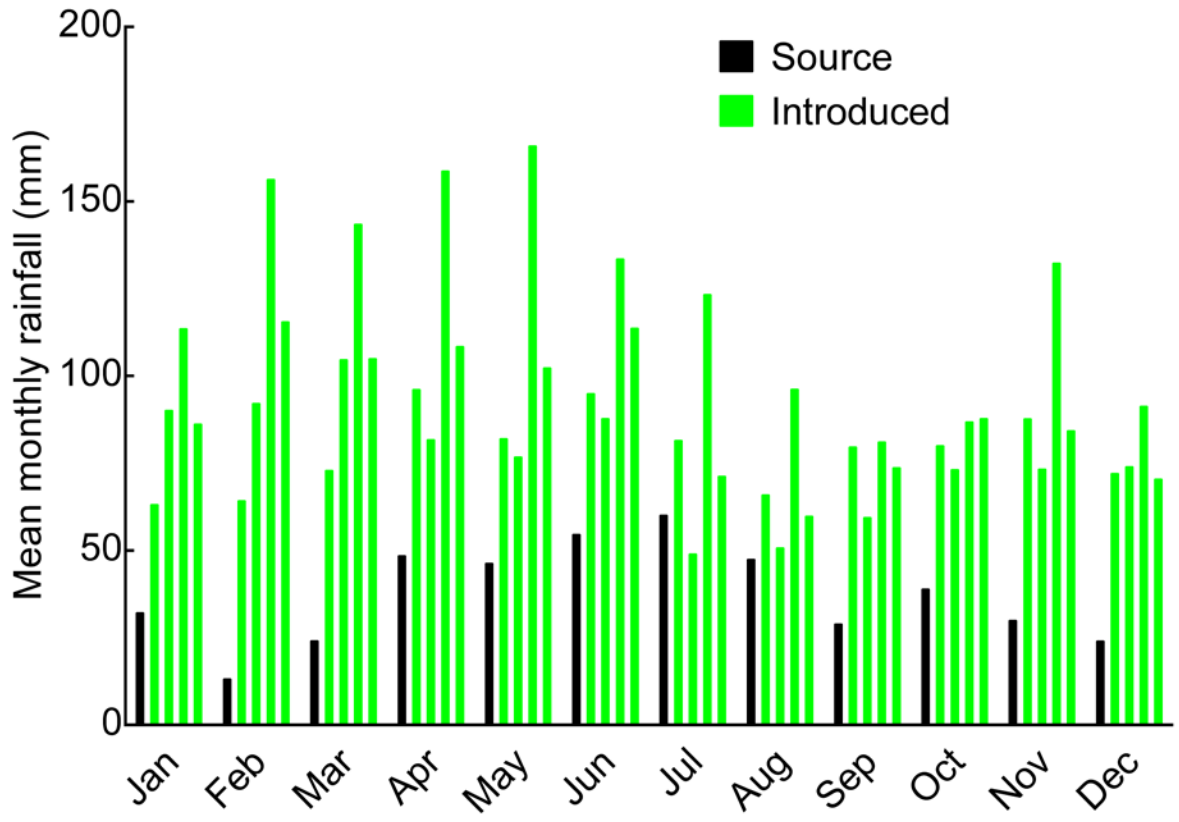
Appendix S1: Location and climate details for sampled populations

Figure S1a. Map showing location of source population (Arniston, South Africa) and the four populations in eastern Australia.

Table S1. GPS co-ordinates of collection locations, and climate data. Climate data obtained with permission from the South African Weather Service and the Australian Government Bureau of Meteorology.

	Latitude	Longitude	Mean minimum temperature (°C)	Mean maximum temperature (°C)	Mean annual rainfall (mm)
Source population (South Africa)					
Arniston	-34.6579	20.2329	13.5	20.4	448
Introduced populations (Australia)					
Treachery beach	-32.4468	152.5202	14.3	22.8	1482
Wairo beach	-35.4423	150.4089	13.1	20.6	1078
Narooma	-36.2238	150.1401	12.0	20.0	912
Mallacoota	-37.5688	149.7621	10.8	19.4	939

Figure S1b. Mean monthly rainfall (mm) experienced by the source population in Arniston, South Africa, and the four introduced populations on the east coast of Australia (ordered from left to right with increasing latitude). Data obtained with permission from the South African Weather Service and the Australian Government Bureau of Meteorology.



Appendix S2. Locating the South African source population for east Australian *Arctotheca populifolia.*

Leaves were collected from 188 plants in ten populations spanning the entire native range of *Arctotheca populifolia,* and from 160 plants in seven populations across Australia [1]. Rollins et al. extracted DNA, and genotyped the plants using seven polymorphic microsatellite loci. Analysis of the microsatellite data showed that there had been two separate introductions of *A. populifolia* to Australia: one to the east coast and one to the west coast [1]. Our focus is on the eastern-Australian introduction.

STRUCTURE and PCA analyses

Rollins et al.'s STRUCTURE [2, 3] analysis revealed that the four eastern-Australian populations of *Arctotheca populifolia* are very similar to one another, and came from western South Africa [1]. STRUCTURE plot membership values for the genetic group holding all eastern Australian plants were highest for plants from Arniston, South Africa (0.975), followed by plants from Muizenberg (0.930) and Cape St Francis (0.838; pers. comm. Rollins). Rollins et al. also presented a principal coordinate analysis (PCA) of the genetic distances between introduced and home range populations of *A. populifolia* [1]. The PCA showed that the Australian populations of *A. populifolia* were most similar to the population from Arniston, with populations from Muizenberg and Cape St Francis next closest. However, STRUCTURE does not provide statistical testing to determine the likely source population for the east-Australian introduction.

Indeed, STRUCTURE amalgamates individuals into groups that appear to be random mating, which is clearly not likely to be the case for populations separated by the Indian Ocean. Thus, although Rollins et al.'s work pointed toward Arniston as a likely source population, further analysis was required. We used two approaches: R_{ST} and *LOD*.

R$_{ST}$ *analysis*

R_{ST} is a population differentiation measure that uses information on not only the allele proportions in the population, but also the length of microsatellite repeat alleles at each locus. R_{ST} is thought to give better analysis of relationship between groups than other differentiation methods that only use proportion data (eg G_{ST}, F_{ST}). However, this better performance occurs under quite restricted conditions, including complete separation of the groups and tens to hundreds of generations since separation of the groups [4]. Fortunately, our system satisfies these conditions: there is little likelihood of gene flow between African and Australian populations, and *A. populifolia* will have been through more than 80 generations since arriving in Australia because it produces seeds within a year of establishing (pers. obs. CRB) and was introduced to Australia in the 1930s [5].

The R_{ST} analysis considered ten locations that were sampled at approximately 200 km intervals along the species' entire range on the coast of Africa [Figure 2a in 1]. These were each compared with pooled data from the same eastern Australian locations shown in Appendix S1. Pooling was justified for two reasons: because of the high genetic similarity and low variability of the

Australian population [1]; and because we were searching for the common origin of all four populations. R_{ST} values were calculated with RstCALC [calculated with RstCALC; 6].

The R_{ST} analysis showed that the east Australian populations were the most similar to the Arniston population (table S2a). However, populations from Rocherpan, Muizenberg, Hartenbos, Cape St Francis and Kenton-on-Sea had R_{ST} values whose standard errors overlapped with Arniston's, so cannot be ruled out as potential source populations based on the R_{ST} analysis. A related differentiation measure, *Delta-mu*^{2} yielded similar results, including placing Arniston as the most similar to the east Australian populations (table S2a).

Table S2a Estimates of RHO and *Delta-mu*^{2}. RHO is the mean estimate of R_{ST} across loci in comparisons between the pooled east Australian *Arctotheca populifolia* populations versus each South African population. *Delta-mu*^{2} is a measure of the differentiation of repeat length between the pooled east Australian *Arctotheca populifolia* populations and each South African population. *Delta-mu*^{2} was calculated using *RSTOUT2a.out*. Standard errors cannot be calculated for this metric. Bold font indicates the South African population most similar to the east Australian population (Arniston, for both metrics).

Population	RHO (standard error)	*Delta-mu*^{2}
Strandfontein	0.648 (0.148)	25.1853
Rocherpan	0.499 (0.130)	10.3839
Muizenberg	0.473 (0.170)	6.1437

Arniston	**0.355 (0.135)**	**3.8767**
Hartenbos	0.531 (0.133)	7.836
Cape St Francis	0.399 (0.109)	6.6081
Kenton-on-Sea	0.499 (0.117)	11.5079
Kei Mouth	0.714 (0.141)	53.4352
Trafalgar	0.980 (0.016)	66.9422
Umlalazi	0.997 (0.002)	67.0621

82

83 ***LOD Analysis***

84 On the basis of the results from R_{ST}, STRUCTURE and PCA analyses, Arniston seemed the most
85 likely source population. However, five populations from Rocherpan, Muizenberg, Hartenbos,
86 Cape St Francis and Kenton-on-Sea are also possible sources, because of the overlap of standard
87 errors in the R_{ST} analysis. Therefore, we compared the probability of Arniston (AR) being the
88 source of the east Australian population with the probability of each of the other five candidate
89 South African populations *'x'* being the source population. In other words, we evaluate two
90 hypotheses: *AR* as source *vs* *'x'* as source. Evaluation was performed using a log odds (LOD)
91 ratio:

92 *Equation 1:* LOD (AR *vs* x) = $log\ (^{P_{AR}}/_{P_x})$

93 where P_{AR} and P_x . are the probabilities of seeing the Australian data if it is a sample from the
94 relevant South African population, either AR, or one of the five competing hypothetical
95 ancestors '*x*'. These probabilities were calculated using the multinomial:

96 $$\textit{Equation 2: } P_{(locus\ A)} = \frac{N!}{Np_{1in\ EA}! * Np_{2\ in\ EA}! * \ldots} * p_1^{Np_{1\ in\ EA1}} * p_2^{Np_{2\ in\ EA}} * \ldots.$$

97 Where *N* is the total number of alleles sampled from Eastern Australia ; $p_1\ p_2$ ... represent each
98 allele found in the east Australian population, with p_1 etc being the allele proportions in the
99 relevant South African population (i.e, AR or its competitor *x*); and $p_{1\ in\ EA}$ etc being the
100 proportions in east Australia. For each South African population, his multinomial is calculated
101 for each locus, then values for each of the seven loci are multiplied together to give P_{AR} or P_x .

102

103 One complication is that all candidate source populations except Arniston were missing at least
104 one of the alleles present in east Australia. This could indicate that a) these populations are not
105 the source, or b) that the alleles were present in the population but were not sampled. We cannot
106 rule the second possibility out, so we included all these populations in the analyses. However, the
107 probabilities (from equation 2) that are used in the LOD are based on allele proportions
108 multiplied across all the alleles present in east Australia. If one of these alleles is zero in the
109 target population, then $P_x = 0$, which makes $log\ (^{P_{AR}}/_{P_x})$ undefined. Note that it is allowable
110 to reverse the comparison and calculate $log\ (^{P_x}/_{P_{AR}})$, but in our case this term will still be
111 undefined when $P_x = 0$. Thus, where an east Australian allele is missing in population '*x*', we
112 must make some alteration. To minimise false discovery of Arniston as the source, we designed
113 the missing allele correction to be conservative (ie. it is biased against showing that Arniston is
114 more likely to be the source). Thus, we applied a correction that makes the allele proportion in
115 population '*x*' more similar to the proportion in east Australia. This was done by adjusting the

116 missing allele in *x* to the highest proportion that could have a 95% chance of not being sampled
117 in a sample of that size:

118 *Equation 3*: $\text{missing allele proportion} = 1 - 10^{[(\log 0.95)/n]}$

119 where *n* is the total number of alleles sampled from *x*. We must then of course reduce all other
120 allele proportions so that allele proportions still sum to unity in population *x*, ie each other allele
121 proportion is multiplied by (1-new 'missing' allele proportion).

122

123 All five of our calculated LOD scores were highly positive (table S1). A positive LOD suggests
124 that Arniston is the likely source, while a negative LOD suggests that the other South African
125 population is the likely source. A LOD score of 2 would indicate that Arniston was 100 times
126 more likely to be the source than the competing (*x*) population, while a LOD score of 3 would
127 indicate that Arniston was 1000 times more likely to be the source than the competing
128 population. It is common for researchers to use thresholds of two or three as cut-offs for
129 assigning source populations [7]. Here, the lowest LOD score was 99.9 (for Cape St Francis;
130 table S1). Another way of assigning statistical significance to LOD scores is to convert the LOD
131 to a chi square value (as recommended by Lander and Krugylak [7]). In this case, all five
132 analyses suggest that Arniston is significantly ($p < 0.0001$, with Bonferroni correction) more
133 likely to be the source population than is any other South African population.

Table S2b. LOD and Chi square estimates for likelihood of Arniston being source of east Australia versus other potential source populations. After Bonferroni correction, all chi square values yield $p < 0.0001$.

Comparison	**LOD**	**Chi square (1df)**
Arniston - Rocherpan	513.0	2362.5
Arniston - Muizenberg	265.6	1223.2
Arniston - Hartenbos	777.2	3579.4
Arniston - Cape St Francis	99.9	459.9
Arniston - Kenton-on-Sea	157.1	723.5

Appendix S3: Additional experimental details

Table S3a. Number of plants from which seeds were collected at each location; number of parent plants planted and number of plants producing seeds in the first year of the experiment (these seeds were then used in the main experiment – in order to minimize any maternal effects).

	Original number of plants collected at each location	Number of parent plants planted in the glasshouse	Number of plants producing seeds
Source population in South Africa			
Arniston	46	143	36
Introduced populations in Australia			
Treachery beach	17	41	20
Wairo beach	38	68	53
Narooma	24	39	26
Mallacoota	45	70	51
Total	**170**	**356**	**186**

Table S3b. Soil composition for the glasshouse experiment.

Soil component	**Amount used**
River sand	20 kg
Cocopeat	25 kg
Osmocote© Exact Standard 5-6 Month fertilizer	65 g
General fertilizer: 26% dolomite; 13% of each of the following: superphosphate, blood and bone, lime, gypsum, potassium nitrate; 9% trace elements (S: 6.29%, Ca: 10%, Mg: 3.6%, Mn: 2.88%, Fe: 2.73%, Cu: 1.25%, Zn: 1%, B: 0.09% and Mo: 0.0038%)	65 g

145 **Table S3c.** Number of plants planted for each population in the main experiment.

	Number of plants for year-long experiment (pot size: 15 cm × 15 cm)	**Number of plants for harvesting at 12 weeks (pot size: diameter 10 cm × height 7 cm)**
Source population		
Arniston, South Africa	123	21
Introduced populations		
Treachery beach	40	13
Wairo beach	68	23
Narooma	39	12
Mallacoota	70	22
Total	**340**	**91**

146

147 **Table S3d.** A test for the effect of maternal line on experimental traits. Results are from a univariate
148 general linear model (GLM) with the trait of interest as a dependent variable and maternal group as a
149 random factor, followed by a Holm-Bonferroni sequential correction. An asterisk denotes a significant
150 effect of maternal line.

	Population				
	Arniston (SA)	Mallacoota (AUS)	Narooma (AUS)	Wairo Beach (AUS)	Treachery Beach (AUS)
Trait	**p-values**				
Plant length	0.088	0.206	0.624	0.281	0.387
Plant height	0.305	0.497	0.464	0.226	0.987
Plant growth form	0.360	0.305	0.639	0.014	<0.001*
Leaf area	0.324	0.060	0.524	0.490	0.200
Leaf shape	0.202	0.188	0.526	0.117	0.229
SLA	<0.001*	0.387	0.412	0.612	0.754
Average leaf thickness	0.001*	0.477	0.227	0.201	0.544
Leaf density	0.257	0.585	0.377	0.885	0.804
LDMC	<0.001*	0.217	0.870	0.430	0.715

151

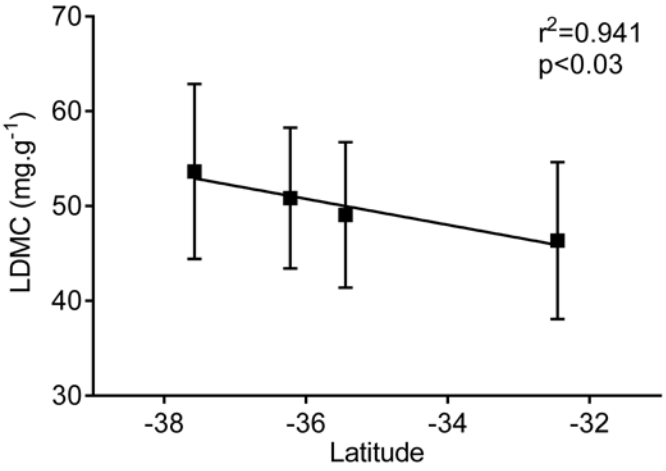
Appendix S4: Differences among introduced populations in Australia

Our main hypothesis concerned the differences between source plants in South Africa and introduced plants in Australia. Nevertheless, we still wanted to test if there were any differences in traits among the four introduced populations in Australia, as it has been shown that introduced plants can evolve clinal differences across the range of their new environment [8]. Leaf dry matter content (LDMC) decreased with increasing latitude ($p<0.03$, figure S4). Leaf density also showed significant differences among Australian populations ($p<0.022$) but since leaf density ≈ leaf dry matter content [9] this correlation is expected. Three other traits (leaf thickness, plant growth form, leaf shape) showed only marginally significant differences ($0.04<p<0.05$) among Australian populations (table S4). When we applied a Holm-Bonferroni sequential correction to counteract the problem of multiple comparisons [10], only LDMC remained significantly different among Australian populations (table S4).

Table S4. Results of one-way analyses of variance (ANOVAs) contrasting each trait among the four introduced populations in Australia. The adjusted p-values are those after a Holm-Bonferroni sequential correction to account for multiple tests [10].

Trait	**MS**	**F**	**p-value**	**Adjusted p-values**
Leaf dry matter content	497.0	7.247	0.001	0.013
Leaf density	515.7	3.265	0.022^	0.264
Leaf thickness	0.028	2.825	0.040	0.440
Plant growth form	0.051	2.714	0.046	0.460
Leaf shape	0.094	2.680	0.048	0.460
Plant length	13651	2.526	0.059	0.472
$\log_{10}$(Specific leaf area)	0.015	2.410	0.068	0.476
Leaf area	105.5	2.087	0.103	0.618
Plant height	3231	1.329	0.266	1.000
Above-ground biomass at 11 months	49.99	0.641	0.593	1.000
Below-ground biomass at 12 weeks	0.019	0.568	0.638	1.000
Total biomass at 12 weeks	0.114	0.480	0.697	1.000
Above-ground biomass at 12 weeks	0.044	0.362	0.780	1.000

^Leaf density ≈ Leaf dry matter content [9]



168

169 **Figure S4.** The correlation between leaf dry matter content (LDMC) and latitude for the four
170 introduced populations in Australia. Points on the graph represent population means (+/- one
171 standard deviation).

Appendix S5: Multivariate analysis of source and introduced plants

Source and introduced plants were found to be significantly different from each other when we tested individuals for differences using a multivariate analysis of variance (MANOVA, $p<0.001$). This kind of analysis uses multiple trait measurements which have been taken on the same individual plant, and so we based our analysis on nine of the twelve traits (biomass measurements at 12 weeks were measured on a separate set of young plants which were harvested; above-ground biomass at senescence was only measured on a subset of plants and therefore did not have a large enough sample size to be included). We created ordination plots using factor analysis [11] which showed that Factors 3 and 4 were more effective at separating out source and introduced plants than Factors 1 and 2 (figure S5). The proportion of variation explained by each trait in the multivariate analysis is given in Table S5.

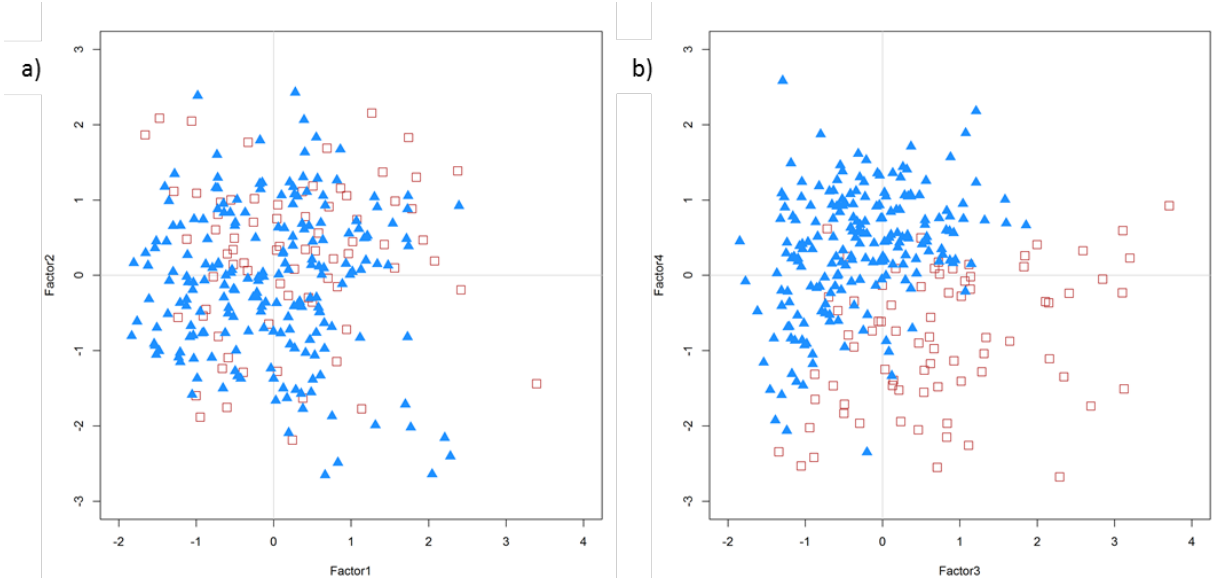


Figure S5. Ordination plots using factor analysis. Source plants are shown as squares and introduced plants are shown as triangles. Panel a) displays the data arrayed on axes of factors 1 and 2, while panel b displays the data arrayed on axes for factors 3 and 4.

187 **Table S5**. The proportion of variation explained by each trait in the multivariate analysis.

Trait	**R^2**
Plant growth form	0.253
Leaf shape	0.230
Plant length	0.226
Leaf area	0.147
Average leaf thickness	0.111
Leaf density	0.073
Plant height	0.041
Leaf dry matter content	0.008
$\log_{10}$(Specific leaf area)	0.007

188

References

1. Rollins LA, Moles AT, Lam S, Buitenwerf R, Buswell JM, Brandenburger CR, Flores-Moreno H, Nielsen KB, Couchman E, Brown GS, et al. 2013 High genetic diversity is not essential for successful introduction. *Ecol Evol* **3**, 4501-4517. (doi:10.1002/ece3.824/pdf).
2. Pritchard JK, Stephens M, Donnelly P. 2000 Inference of population structure using multilocus genotype data. *Genetics* **155**, 945-959.
3. Falush D, Stephens M, Pritchard JK. 2003 Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. *Genetics* **164**, 1567-1587.
4. Balloux F, Lugon-Moulin N. 2002 The estimation of population differentiation with microsatellite markers. *Molecular ecology* **11**, 155-165. (doi:10.1046/j.0962-1083.2001.01436.x).
5. AVH Database. 2016 Australia's Virtual Herbarium, Council of Heads of Australasian Herbaria. http://avh.ala.org.au. Accessed on 10 August 2016.
6. Goodman SJ. 1997 RST Calc: a collection of computer programs for calculating estimates of genetic differentiation from microsatellite data and determining their significance. *Molecular Ecology* **6**, 881-885. (doi:10.1046/j.1365-294X.1997.00260.x).
7. Lander E, Kruglyak L. 1995 Genetic dissection of complex traits: guidelines for interpreting and reporting linkage results. *Nat Genet* **11**, 241. (doi:10.1038/ng1195-241).
8. Maron JL, Vilà M, Bommarco R, Elmendorf S, Beardsley P. 2004 Rapid evolution of an invasive plant. *Ecol Monogr* **74**, 261-280. (doi:10.1890/03-4027).

9. Pérez-Harguindeguy N, Díaz S, Garnier E, Lavorel S, Poorter H, Jaureguiberry P, Bret-Harte M, Cornwell W, Craine J, Gurvich D. 2013 New handbook for standardised measurement of plant functional traits worldwide. *Aust J Bot* **61**, 167-234. (doi:10.1071/BT12225).

10. Holm S. 1979 A simple sequentially rejective multiple test procedure. *Scand J Stat*, 65-70.

11. Everett B. 2013 *An introduction to latent variable models*, Springer Science & Business Media.