Supplementary Information

Genetic patterns and climate modelling reveal challenges for conserving *Sclerolaena napiformis* (Amaranthaceae s.l.) an endemic chenopod of southeastern Australia

Michael Amor [1]*, Neville Walsh [1] and Elizabeth James [1]

[1] Royal Botanic Gardens Victoria; michael.amor@rbg.vic.gov.au
[1] Royal Botanic Gardens Victoria; neville.walsh@rbg.vic.gov.au
[1] Royal Botanic Gardens Victoria; elizabeth.james@rbg.vic.gov.au
* Correspondence: michael.amor@rbg.vic.gov.au

Table S1: Environmental and soil variables used for modelling suitable habitat for *Sclerolaena napiformis*. Bioclimatic variables (30 seconds resolution) were downloaded from the WorldClim database (www.worldclim.org/bioclim). Soil variables were downloaded from the CSIRO soil and landscape database (http://www.clw.csiro.au/aclep/soilandlandscapegrid/index.html).

Table S2: Analysis of molecular variance among disjunct southwest, central and northeast regions of *Sclerolaena napiformis*. Analysis was validated by 1000 permutations.

Table S3: Analysis of molecular variance among southwestern *Sclerolaena napiformis* sites. Analysis was validated by 1000 permutations.

Table S4: Analysis of molecular variance among central *Sclerolaena napiformis* sites. Analysis includes investigations of variance (a) between Victoria and New South Wales (states are divided by the Murray River), and (b) among all sites without consideration of state variable. Analysis was validated by 1000 permutations.

Table S5: Analysis of molecular variance among northeast *Sclerolaena napiformis* sites. Analysis was validated by 1000 permutations.

Table S6: Lower, mean and upper F_{IS} estimates for each collection site of *Sclerolaena napiformis*. Calculations are based on 4,000+ ddRADseq loci from regional assemblies (see Table 2 of the manuscript).

Figure S1: a) Receiver Operating Characteristic (ROC) curve, estimating the accuracy of our model based on presence/absence data. Estimates were performed by randomly simulating 1,000 pseudoabsences within our sampling area of extent which was tested against our sampling observations/collections. We show high confidence that our sampling is representative of the population; b) Percentage contributions of soil and climate variables when modelling the distribution of *Sclerolaena napiformis* based on our occurrence data. BioClim climate data were obtained from the WorldClim database. Soil data were obtained from the CSIRO Soil and Landscape Grid.

Figure S2: DAPC outputs from based on analysis of our 'all regions' assembly (southwest, central and northeast) showing BIC based estimates of cluster numbers, corresponding assignment probabilities per individual and membership assignments per group.

Figure S3 DAPC outputs from based on analysis of our combined 'central' and 'northeast' assembly showing (a) BIC based estimates of cluster numbers, corresponding DAPC plots for (b) K=4 and (c) K=7, group assignments for (d) K=4 and (e) K=7, membership assignment probabilities for (f) K=4 and (g) K=7. FastStructure (h) Delta K and (i) probability plots generated via Clumpak are also shown.

Figure S4 DAPC outputs from based on analysis of our 'southwest' assembly showing (a) BIC based estimates of cluster numbers, corresponding DAPC plots for (b) K=6 and (c) K=8 (d) K=9, (e) K=10, group assignments for (f) K=6, (g) K=8 (h) K=9, (i) K=10. FastStructure (j) Delta K and (k) probability plots generated via Clumpak are also shown.

Figure S5. Haplotype network and distribution of haplotypes based on mapping of ddRADseq loci to chloroplast reference genomes. a) The haplotype network consisted of a single maternal lineage dominated by one haplotype (H1) that was found in 351/390 individuals. Haplotypes 2 – 6 were found in 21, 5, 9, 3 and 1 individual/s, respectively. b) geographic distribution of haplotypes across sites ordered west to east. H1 was found in the highest proportion at every site, except site 25, with no evidence of geographic patterns.

Table S1: Environmental and soil variables used for modelling suitable habitat for *Sclerolaena napiformis*. Bioclimatic variables (30 seconds resolution) were downloaded from the WorldClim database (www.worldclim.org/bioclim). Soil variables were downloaded from the CSIRO soil and landscape database (http://www.clw.csiro.au/aclep/soilandlandscapegrid/index.html).

Variable	Details
BIO1	Annual Mean Temperature
BIO5	Max Temperature of Warmest Month
BIO6	Min Temperature of Coldest Month
BIO7	Temperature Annual Range (BIO5-BIO6)
BIO8	Mean Temperature of Wettest Quarter
BIO9	Mean Temperature of Driest Quarter
BIO12	Annual Precipitation
BIO15	Precipitation Seasonality (Coefficient of Variation)
BIO16	Precipitation of Wettest Quarter
BIO17	Precipitation of Driest Quarter
BIO18	Precipitation of Warmest Quarter
BIO19	Precipitation of Coldest Quarter
Clay	< 2 um mass fraction of the < 2 mm soil material
Bulk Density - Whole Earth	Bulk Density of the whole soil (including coarse fragments)
Organic Carbon	mass fraction of carbon by weight in the < 2 mm soil material
Silt	2-20 um mass fraction of the < 2 mm soil material
Sand	20 um - 2 mm mass fraction of the < 2 mm soil material
Total Nitrogen	Total nitrogen
Total Phosphorus	Total phosphorus
Available Water Capacity	Available water capacity

Table S2: Analysis of molecular variance among disjunct southwest, central and northeast regions of *Sclerolaena napiformis*. Analysis was validated by 1000 permutations.

	SSD	MSD	df	sigma2	Variation (%)	P value
Regional analysis	190.927	95.464	2	1.338	8.697	0.000
Error	3552.253	14.041	253	14.041		
Total	3743.180	14.679	255			

Table S3: Analysis of molecular variance among southwestern *Sclerolaena napiformis* sites. Analysis was validated by 1000 permutations.

	SSD	MSD	df	sigma2	Variation (%)	P value
Among sites	23.263	23.263	1	0.123	0.910	0.011
Error	2270.971	13.438	169	13.438	99.090	
Total	2294.234	13.495	170			

Table S4: Analysis of molecular variance among central *Sclerolaena napiformis* sites. Analysis includes investigations of variance (a) between Victoria and New South Wales (states are divided by the Murray River), and (b) among all sites without consideration of state variable. Analysis was validated by 1000 permutations.

	SSD	MSD	df	sigma2	Variation (%)	P value
Between VIC/NSW	9.913	9.913	1	0.119	2.338	0.0250
Among sites	29.094	5.819	5	0.092	1.806	0.0829
Error	308.465	4.896	63	4.896	95.856	
Total	347.472	5.036	69			

Table S5: Analysis of molecular variance among northeast *Sclerolaena napiformis* sites. Analysis was validated by 1000 permutations.

	SSD	MSD	df	sigma2	Variation (%)	P value
Among sites	19.627	19.627	1	1.374	11.603	0.014
Error	136.077	10.467	13	10.468		
Total	155.704	11.122	14			

Table S6: Lower, mean and upper F_{IS} estimates for each collection site of *Sclerolaena napiformis*. Calculations are based on 4,000+ ddRADseq loci from regional assemblies (see Table 2 of the manuscript).

Site	Region / assembly	n	Fis Lower (0.025)	Fis (Mean)	Fis Upper (0.975)
SN15	Central	18	0.149	0.244	0.339
SN16	Central	14	0.332	0.416	0.500
SN17	Central	20	0.693	0.795	0.897
SN18	Central	20	0.187	0.344	0.501
SN21	Central	14	0.511	0.565	0.619
SN22	Central	20	0.418	0.498	0.578
SN23	Central	20	0.539	0.615	0.691
SN19	Northeast	8	0.737	0.776	0.815
SN20	Northeast	22	0.776	0.811	0.846
SN01	Southwest	8	0.802	0.856	0.910
SN02	Southwest	19	0.252	0.313	0.374
SN03	Southwest	20	0.507	0.675	0.843
SN04	Southwest	20	0.436	0.542	0.648
SN05	Southwest	19	0.467	0.538	0.609
SN06	Southwest	14	0.502	0.586	0.669
SN07	Southwest	20	0.372	0.423	0.474
SN08	Southwest	15	0.420	0.479	0.538
SN09	Southwest	14	0.333	0.406	0.480
SN10	Southwest	16	0.191	0.250	0.309
SN11	Southwest	16	0.463	0.529	0.595
SN12	Southwest	21	0.349	0.403	0.456
SN13	Southwest	19	0.075	0.125	0.176
SN14	Southwest	14	0.517	0.568	0.618
SN24	Southwest	6	-0.991	-0.539	-0.088
SN25	Southwest	20	0.705	0.782	0.858
SN26	Southwest	19	0.351	0.455	0.559
SN27	Southwest	15	0.475	0.533	0.592

Figure S1: a) Receiver Operating Characteristic (ROC) curve, estimating the accuracy of our model based on presence/absence data. Estimates were performed by randomly simulating 1,000 pseudoabsences within our sampling area of extent which was tested against our sampling observations/collections. We show high confidence that our sampling is representative of the population; b) Percentage contributions of soil and climate variables when modelling the distribution of *Sclerolaena napiformis* based on our occurrence data. BioClim climate data were obtained from the WorldClim database. Soil data were obtained from the CSIRO Soil and Landscape Grid.

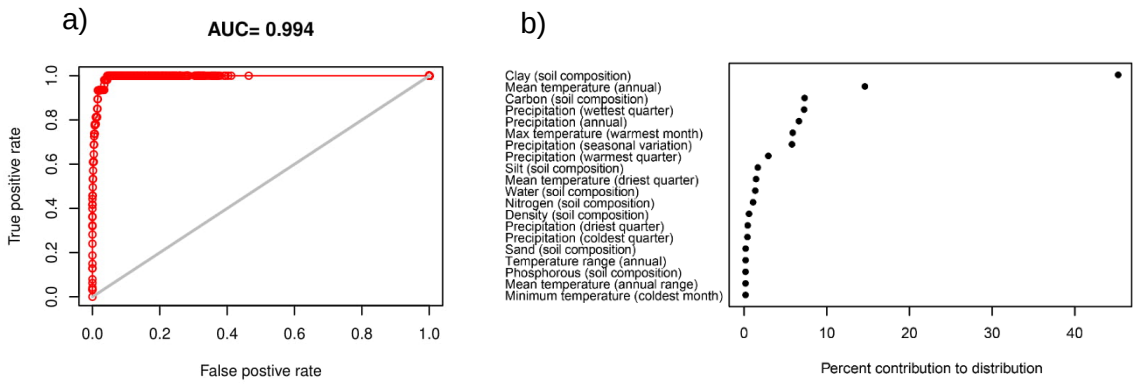


Figure S2: DAPC outputs from based on analysis of our 'all regions' assembly (southwest, central and northeast) showing BIC based estimates of cluster numbers, corresponding assignment probabilities per individual and membership assignments per group.

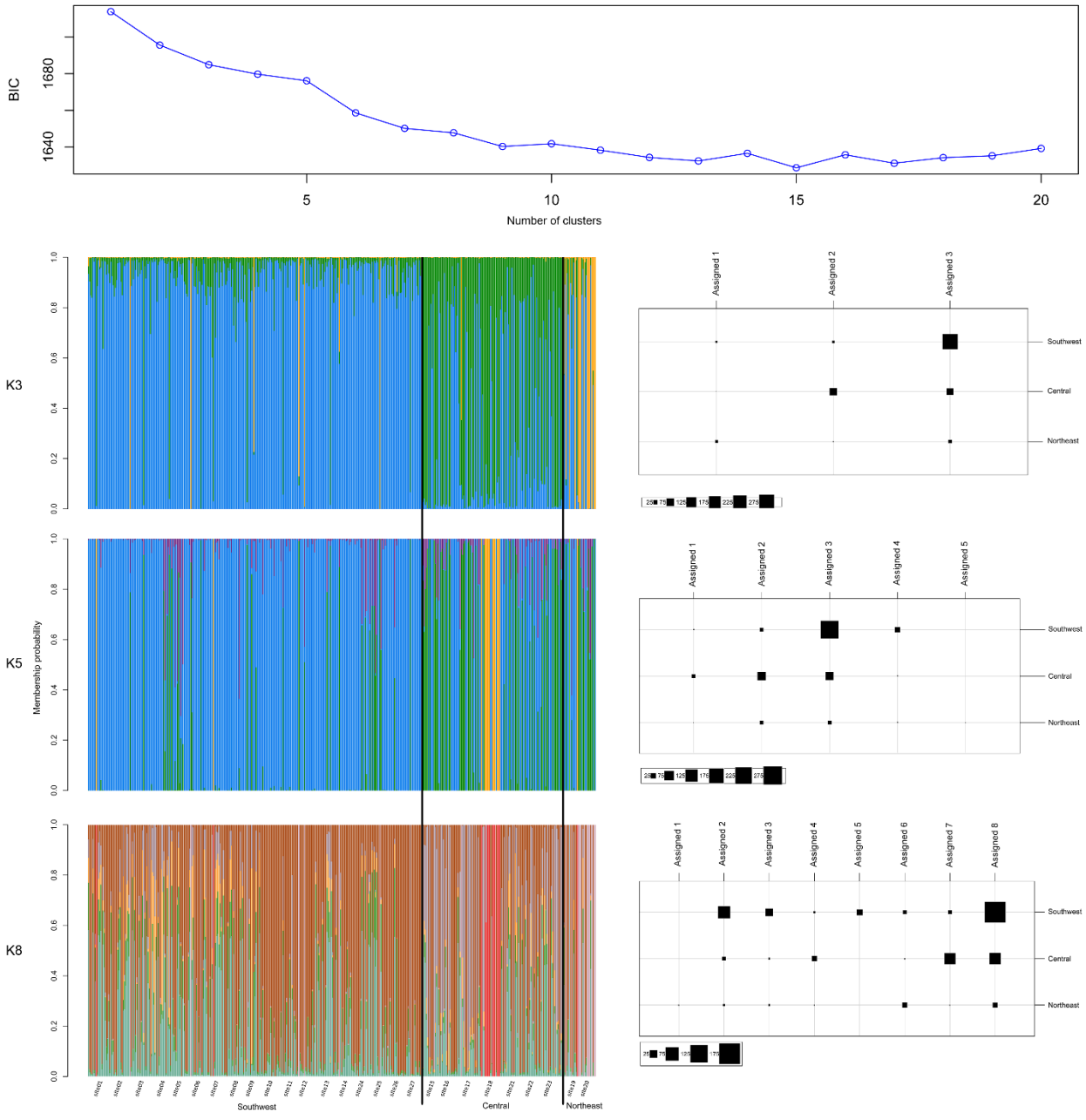
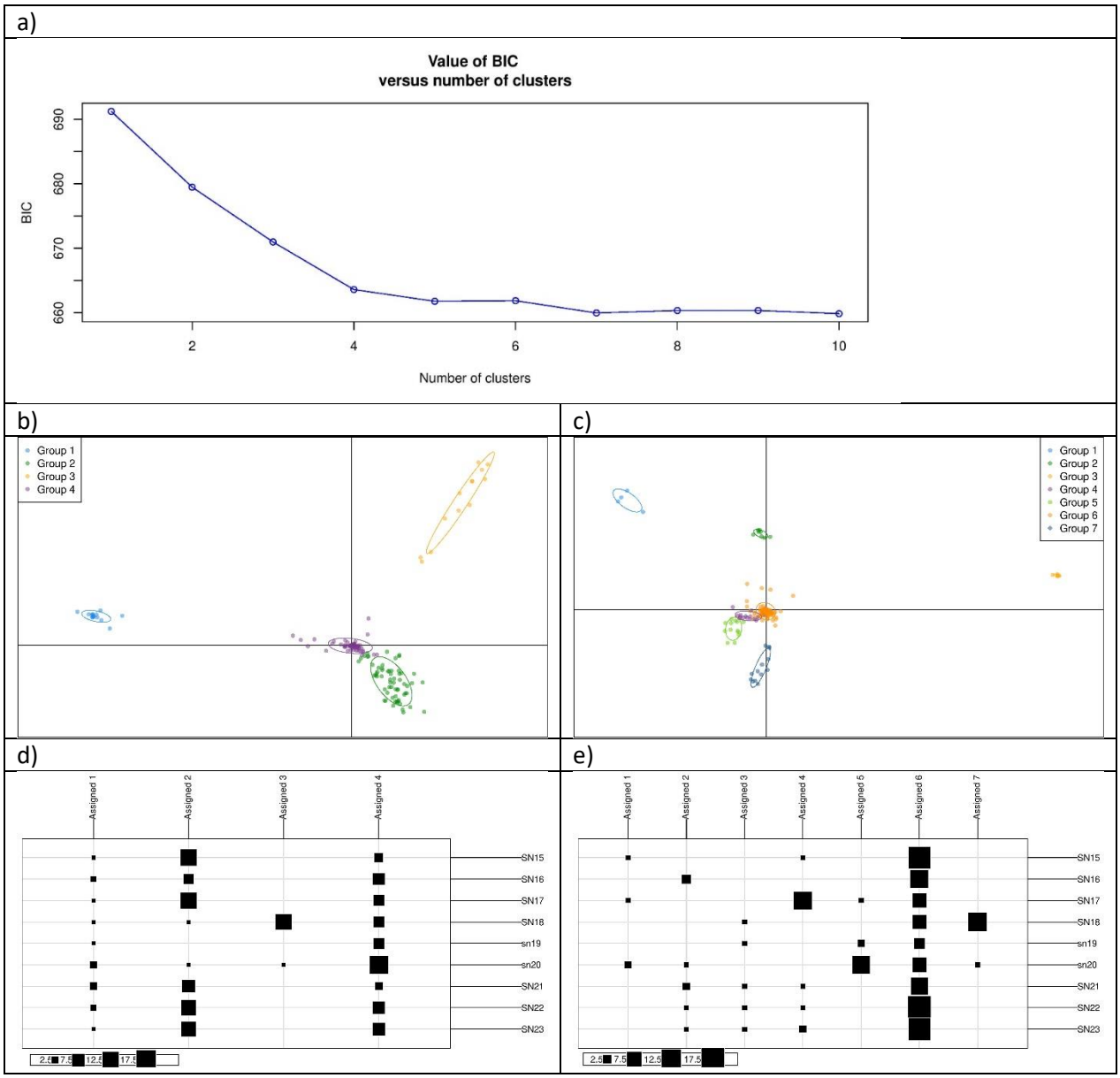
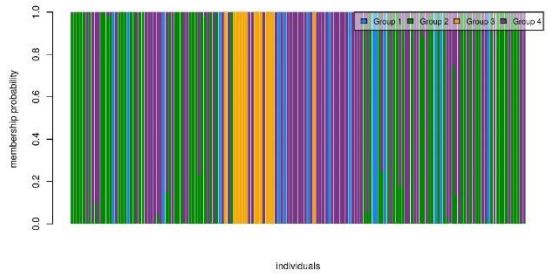


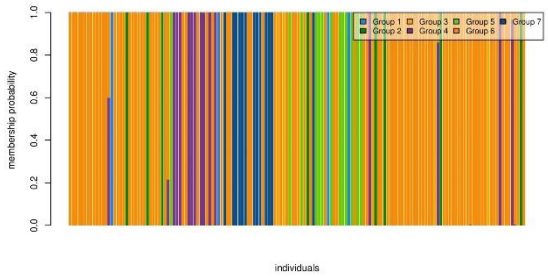
Figure S3 DAPC outputs from based on analysis of our combined 'central' and 'northeast' assembly showing (a) BIC based estimates of cluster numbers, corresponding DAPC plots for (b) K=4 and (c) K=7, group assignments for (d) K=4 and (e) K=7, membership assignment probabilities for (f) K=4 and (g) K=7. FastStructure (h) Delta K and (i) probability plots generated via Clumpak are also shown.



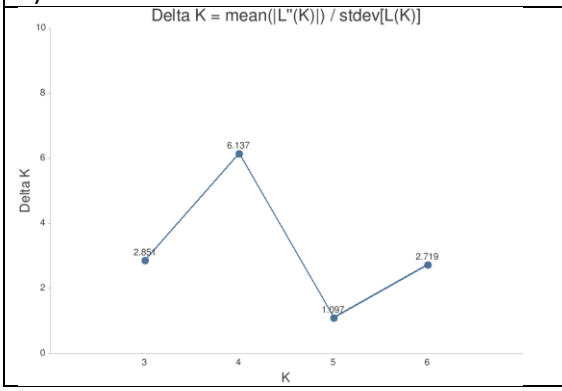
f)


Group 1
Group 2
Group 3
Group 4
membership probability
0.0
0.2
0.4
0.6
0.8
1.0
individuals

g)


Group 1
Group 2
Group 3
Group 4
Group 5
Group 6
Group 7
membership probability
0.0
0.2
0.4
0.6
0.8
1.0
individuals

h)


Delta K = mean(|L''(K)|) / stdev[L(K)]
Delta K
0
2
4
6
8
10
2.851
6.137
1.097
2.719
3
4
5
6
K

i)

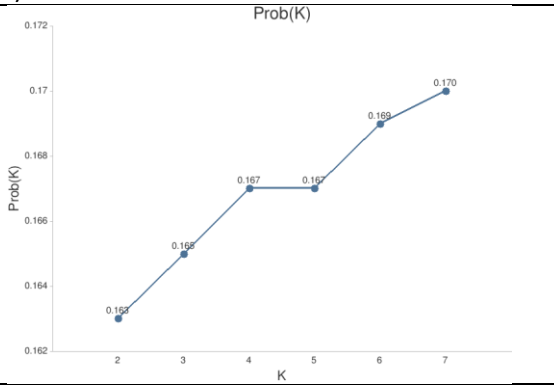
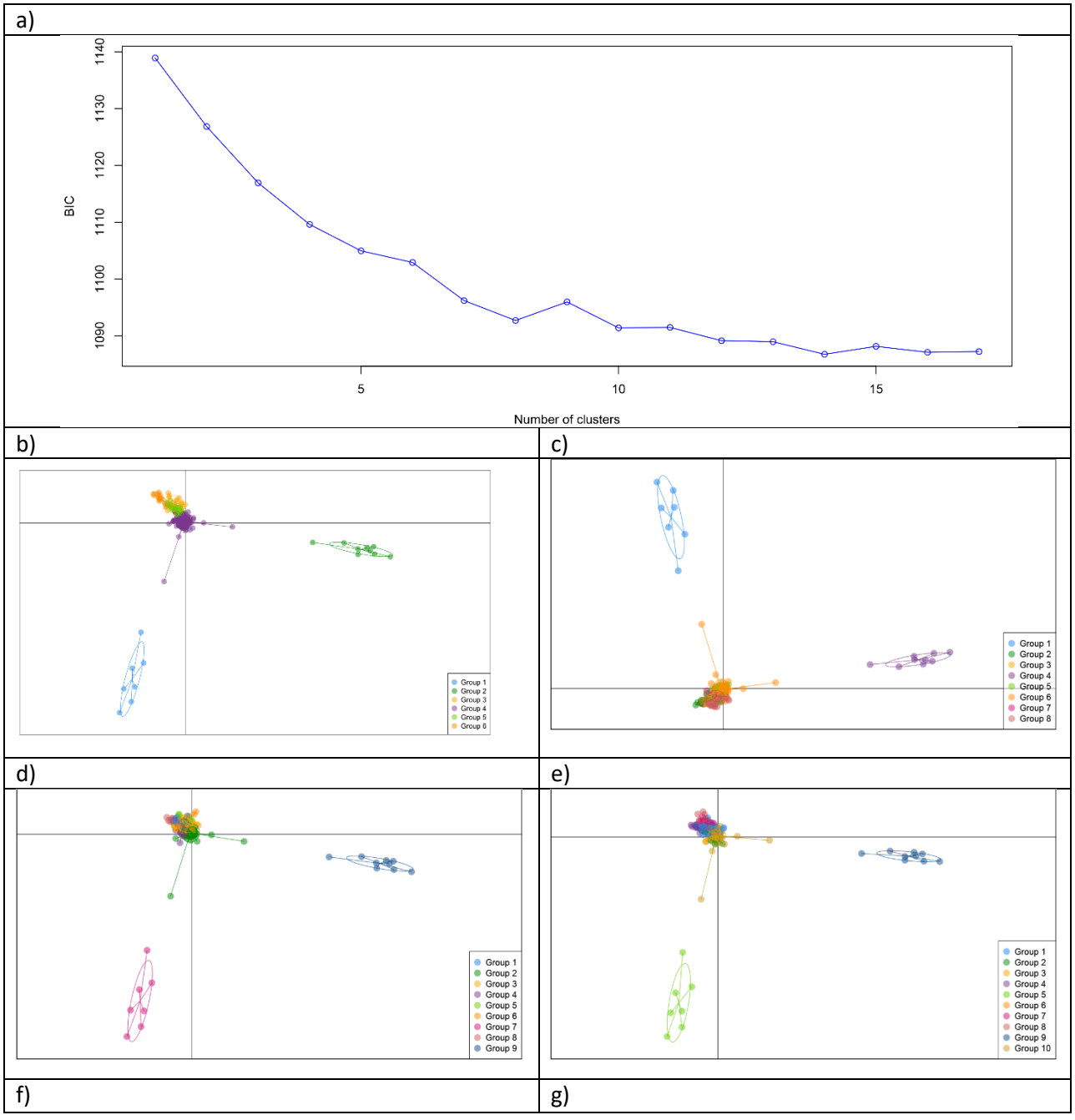
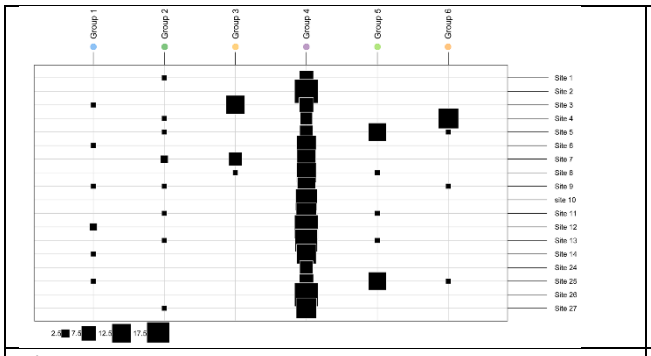
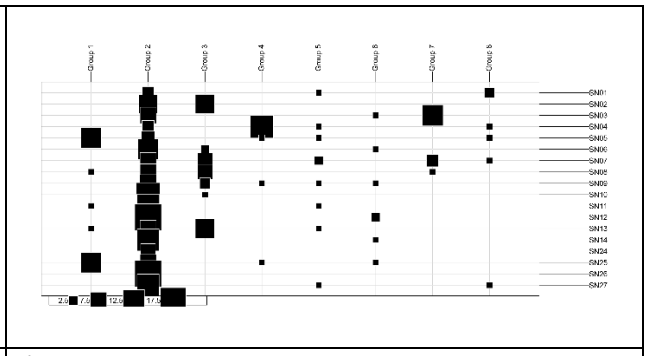

Prob(K)
Prob(K)
0.162
0.164
0.166
0.168
0.17
0.172
0.163
0.168
0.167
0.167
0.169
0.170
2
3
4
5
6
7
K

Figure S4 DAPC outputs from based on analysis of our 'southwest' assembly showing (a) BIC based estimates of cluster numbers, corresponding DAPC plots for (b) K=6 and (c) K=8 (d) K=9, (e) K=10, group assignments for (f) K=6, (g) K=8 (h) K=9, (i) K=10. FastStructure (j) Delta K and (k) probability plots generated via Clumpak are also shown.

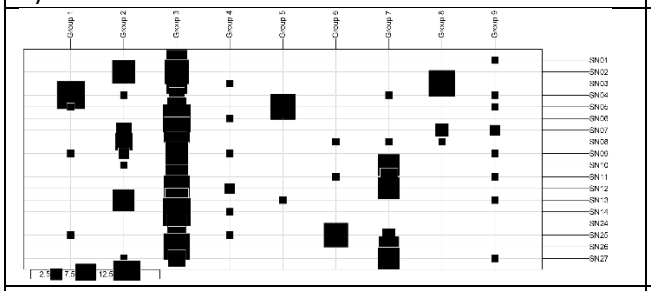



Group 1
Group 2
Group 3
Group 4
Group 5
Group 6
Site 1
Site 2
Site 3
Site 4
Site 5
Site 6
Site 7
Site 8
Site 9
site 10
Site 11
Site 12
Site 13
Site 14
Site 24
Site 25
Site 26
Site 27

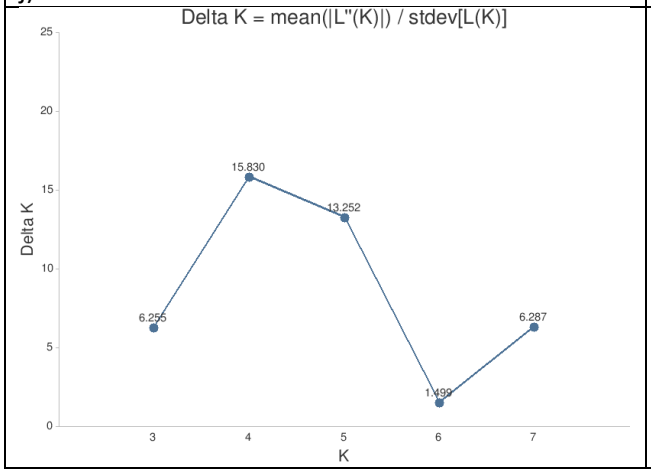
h)



i)



j)


Delta K = mean(|L"(K)|) / stdev[L(K)]
Delta K
K
15.830
13.252
6.255
1.492
6.287

k)

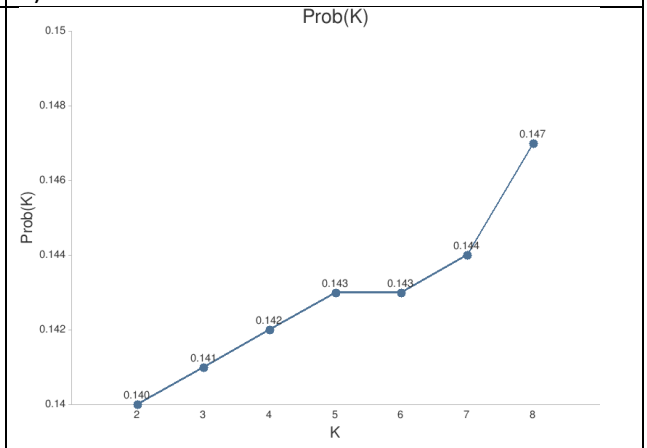

Prob(K)
Prob(K)
K
0.140
0.141
0.142
0.143
0.143
0.144
0.147

Figure S5. Haplotype network and distribution of haplotypes based on mapping of ddRADseq loci to chloroplast reference genomes. a) The haplotype network consisted of a single maternal lineage dominated by one haplotype (H1) that was found in 351/390 individuals. Haplotypes 2 – 6 were found in 21, 5, 9, 3 and 1 individual/s, respectively. b) geographic distribution of haplotypes across sites ordered west to east. H1 was found in the highest proportion at every site, except site 25, with no evidence of geographic patterns.